

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 45.463 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKLHHHHHTSAGITR.....TTWTSAMWHPQGGKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1771	100.0	334	5	ABG32182 HCV prote
2	1660	93.7	409	5	ABG32181 HCV prote
3	1589	89.7	303	5	ABG32183 HCV prote
4	1589	89.7	341	5	ABG32187 HCV prote
5	1589	89.7	352	5	ABG32186 HCV prote
6	1589	89.7	380	5	ABG32185 HCV prote
7	1589	89.7	393	5	ABG32184 HCV prote
8	1580	89.2	303	5	ABG32191 HCV prote
9	1579	89.2	303	5	ABG32189 HCV prote
10	1570	88.7	301	5	ABG32190 HCV prote
11	1531	86.5	292	5	ABG32188 HCV prote
12	1531	86.4	2201	5	ABG30591 Hepatitis
13	1531	86.4	2201	5	ABG30592 Hepatitis
14	1531	86.4	2201	5	ABG30593 Hepatitis
15	1531	86.4	2201	5	ABG30581 Hepatitis
16	1531	86.4	2201	5	ABG30582 Hepatitis
17	1531	86.4	2201	5	ABG30583 Hepatitis
18	1531	86.4	2201	5	ABG30580 Hepatitis
19	1531	86.4	2201	5	ABG30587 Hepatitis
20	1531	86.4	2201	5	ABG30599 Hepatitis
21	1531	86.4	2201	5	ABG30594 Hepatitis
22	1531	86.4	2201	5	ABG30598 Hepatitis
23	1531	86.4	2201	5	ABG30595 Hepatitis
24	1531	86.4	3010	5	ABG32458 Hepatitis
25	1531	86.4	3010	5	ABG32459 Hepatitis

26	1531	86.4	3010	5	ABG32451 Hepatitis
27	1531	86.4	3010	5	ABG32455 Hepatitis
28	1531	86.4	3010	5	ABG32457 Hepatitis
29	1531	86.4	3010	5	ABG32460 Hepatitis
30	1531	86.4	3010	5	ABG32461 Hepatitis
31	1531	86.4	3010	5	ABG32454 Hepatitis
32	1531	86.4	3010	5	ABG32456 Hepatitis
33	1531	86.4	3011	5	ABG32457 HCV-S1 fu
34	1528	86.3	2201	5	ABG30586 Hepatitis
35	1528	86.3	2201	5	ABG30589 Hepatitis
36	1528	86.3	2201	5	ABG30583 Hepatitis
37	1528	86.3	2201	5	ABG30588 Hepatitis
38	1528	86.3	2307	3	AA70064 Recombina
39	1528	86.3	3010	2	AA68622 HCV prote
40	1528	86.3	3010	2	AA68622 HCV prote
41	1527	86.2	2201	5	ABG30590 Hepatitis
42	1525	86.1	2307	3	AA70065 Recombina
43	1525	86.1	3010	5	ABG32452 Hepatitis
44	1524	86.1	2201	5	ABG30584 Hepatitis
45	1524	86.1	2201	5	ABG30602 Hepatitis

## ALIGNMENTS

RESULT 1	
ABG32182	
ID	ABG32182 standard; protein; 334 AA.
AC	ABG32182;
XX	
DT	05-NOV-2002 (first entry)
XX	
DE	HCV protease NS2/3 truncation 4K-6H (904-1206) st-4K.
XX	
KM	HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
KM	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
KM	hepatotropic; antiinflammatory; leucylalanylamine oxide; LDMO;
KM	chaotropic agent; 4K-6H (904-1206) st-4K; mutant; mutain.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
XX	
FH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..15
FT	/note="4-Lys/His tag"
FT	Protein
FT	16..302
FT	/note="Truncated NS2/3 protease"
FT	Peptide
FT	319..334
FT	/note="Streptavidin/4-Lys tag"
XX	
WO	WO200248375-A2.
XX	
PD	20-JUN-2002.
XX	
PF	13-DEC-2001; 2001MO-CA001796.
XX	
PR	15-DEC-2000; 2000US-0256031P.
XX	
PA	(BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX	
PI	Thibeault D, Lamarrre D, Maurice R, Pilote L, Pause A;
XX	
DR	WPI; 2002-599511/64.
XX	
XX	N-PSDB; ABR90407.
PT	Novel polypeptide for screening inhibitors of non-structural proteases
PT	useful as therapeutic agents against hepatitis C virus, comprises full
XX	length non-structural protease, or its truncation.
XX	
PS	Claim 39; Fig 9B; 67bp; English.
XX	
CC	The invention relates to an isolated polypeptide consisting of a full-

CC Length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32181; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent, and LDAO in the presence of reduced concentration of the  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 4K-5H (904-  
 CC 1206)/st-4K comprising a truncated NS2/3 protein with a four Lys/six His N  
 CC -terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag  
 XX

SO Sequence 334 AA:

Query Match 100.0%; Score 1771; DB 5; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 5, 4e-164;  
 Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKKKKLEHHHHHTSAGITKPYFVRAQGLIRACMLVRKAGSHYQMAFMKLAALTGY 60  
 Db 1 MKKKKLEHHHHHTSAGITKPYFVRAQGLIRACMLVRKAGSHYQMAFMKLAALTGY 60  
 61 VYDHLTFLQDMANAGLDLAVAVEPVFSDMEVKIITWGDITACGDIISGLPVSARRR 120  
 Qy 61 VYDHLTFLQDMANAGLDLAVAVEPVFSDMEVKIITWGDITACGDIISGLPVSARRR 120  
 121 EILGPRDNEGGGWRLLAPITAYSOOTRGLGCIITSLGRKXNVEGVEQVVFATOS 180  
 Qy 121 EILGPRDNEGGGWRLLAPITAYSOOTRGLGCIITSLGRKXNVEGVEQVVFATOS 180  
 121 EILGPRDNEGGGWRLLAPITAYSOOTRGLGCIITSLGRKXNVEGVEQVVFATOS 180  
 Db 121 EILGPRDNEGGGWRLLAPITAYSOOTRGLGCIITSLGRKXNVEGVEQVVFATOS 180  
 181 FLATCVAGCMVTFHAGSKTLGAPKPIITOMYTNDODLVGQAPPGARSMTPTCGSS 240  
 Qy 181 FLATCVAGCMVTFHAGSKTLGAPKPIITOMYTNDODLVGQAPPGARSMTPTCGSS 240  
 181 FLATCVAGCMVTFHAGSKTLGAPKPIITOMYTNDODLVGQAPPGARSMTPTCGSS 240  
 Db 181 FLATCVAGCMVTFHAGSKTLGAPKPIITOMYTNDODLVGQAPPGARSMTPTCGSS 240  
 241 DLYIVTRHADVTPRRRGDSRGLSLSPRVSYSKGGSGCGLLPPSHAVGIFPAACVTS 300  
 Qy 241 DLYIVTRHADVTPRRRGDSRGLSLSPRVSYSKGGSGCGLLPPSHAVGIFPAACVTS 300  
 241 DLYIVTRHADVTPRRRGDSRGLSLSPRVSYSKGGSGCGLLPPSHAVGIFPAACVTS 300  
 Db 241 DLYIVTRHADVTPRRRGDSRGLSLSPRVSYSKGGSGCGLLPPSHAVGIFPAACVTS 300  
 301 VAKAVDFIPVESMETTWTSSAMRHPPFGGKXXX 334  
 Qy 301 VAKAVDFIPVESMETTWTSSAMRHPPFGGKXXX 334  
 301 VAKAVDFIPVESMETTWTSSAMRHPPFGGKXXX 334  
 Db 301 VAKAVDFIPVESMETTWTSSAMRHPPFGGKXXX 334

RESULT 2  
 ABG32181  
 ID ABG32181 standard; protein; 409 AA.

XX AC ABG32181;  
 XX DT 05-NOV-2002 (first entry)  
 XX

DE HCV protease NS2/3 (810-1206).  
 XX HCV, enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key location/Qualifiers  
 FT Peptide 398, 409  
 FT /note= "Streptavidin tag"  
 XX  
 XX WO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-COA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;  
 XX WPI, 2002-599511/64.  
 XX DR N-PSDB; ABR90406.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 PS Claim 42; Fig 1B; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32181; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 (810-1206) protein, which has a C-  
 CC terminal streptavidin tag  
 XX

Query Match 93.7%; Score 1660; DB 5; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 4, 9e-153;



OS Synthetic.  
XX WO200248375-A2.  
XX  
XX 20-JUN-2002.  
PD  
XX 13-DEC-2001; 2001WO-CA001796.  
PF  
XX 15-DEC-2000; 2000US-0256031P.  
PR  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
PA  
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
PI  
XX WPI; 2002-599511/64.  
DR  
XX  
XX  
PT Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
PS  
XX Claim 41; Page 62-63; 67pp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32186; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence of  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC numbered sequence represents the NS2/3 truncation mutant 866-1206  
CC (numbered relative to the full length NS2/3 protein)  
XX  
SQ Sequence 341 AA;  
Query March 89.7%; Score 1589; DB 5; Length 341;  
Best Local Similarity 100.0%; Pred. No. 3.2e-146;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 16 AGTTKVVYFPAAGLIRACLVKRAAGGVVQVAFMKLALTTGYVDHLPQDMAHAG 75  
Db 39 AGTTKVVYFPAAGLIRACLVKRAAGGVVQVAFMKLALTTGYVDHLPQDMAHAG 98  
QY 76 LRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGQGM 135  
Db 99 LRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRGREILIGPADNFGQGM 158  
QY 136 RLAPITAVSQGTGRLIGCTITSLTGDRKNQVGEVGVSTAQSFATCVCVCTVTR 195  
Db 159 RLAPITAVSQGTGRLIGCTITSLTGDRKNQVGEVGVSTAQSFATCVCVCTVTR 218

QY 196 GAGSKTLAEPKGPITTYNTVDOLVGMQAPRGARSNTPECTCGSSDLYLVTTRADVTPVR 255  
Db 219 GAGSKTLAEPKGPITTYNTVDOLVGMQAPRGARSNTPECTCGSSDLYLVTTRADVTPVR 278  
QY 256 RRGDSRGSILSPREVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
Db 279 RRGDSRGSILSPREVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338  
QY 316 TMR 318  
Db 339 TMR 341  
RESULT 5  
ABG32186  
ID ABG32186 standard; protein; 352 AA.  
XX  
XX ABG32186;  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX  
XX HCV protease NS2/3 truncation mutant 855-1206.  
XX  
XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
XX chaotropic agent; mutant; mucin.  
XX  
XX Hepatitis C virus.  
XX Synthetic.  
XX  
XX WO200248375-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 13-DEC-2001; 2001WO-CA001796.  
XX  
XX 15-DEC-2000; 2000US-0256031P.  
XX  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX  
XX Claim 41; Page 61-62; 67pp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32186; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence of



Query Match	89.7%	Score 1589;	DB 5;	Length 352;
Best Local Similarity	100.0%	Pred. No. 3.4e-146;		
Matches 303;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	16	AGITKVPFVFAOGLIRACMLVTKRAAGGHVYQMAFMKLAALTGTVYDHLTEPLQDMANAG	75	
Db	50	AGITKVPFVFAOGLIRACMLVTKRAAGGHVYQMAFMKLAALTGTVYDHLTEPLQDMANAG	109	
QY	76	LRDLAAVNEPVYFISMEVYKIIITWGDPTAACGIIISGLPVSARGREILGPDNNEGQW	135	
Db	110	LRDLAAVNEPVYFISMEVYKIIITWGDPTAACGIIISGLPVSARGREILGPDNNEGQW	169	
QY	136	RLAPITVYSOOTRGLGCIITSLGRDKNQVEGEVQVSTATQSFATCNVGCWTVFH	195	
Db	170	RLAPITVYSOOTRGLGCIITSLGRDKNQVEGEVQVSTATQSFATCNVGCWTVFH	229	
QY	196	GAGSKTLAGPKPITQMTYTNVDQDILVQWAPPGARSMPTCTCGSSDLYVTFHADYIPVR	255	
Db	230	GAGSKTLAGPKPITQMTYTNVDQDILVQWAPPGARSMPTCTCGSSDLYVTFHADYIPVR	289	
QY	256	RRGDSRGSLLSPRPVSYLKGGSSGGPLTQSGHAVGIFRAVCTCRGAKAVDFIPVSMET	315	
Db	290	RRGDSRGSLLSPRPVSYLKGGSSGGPLTQSGHAVGIFRAVCTCRGAKAVDFIPVSMET	349	
QY	316	TMR 318		
Db	350	TMR 352		
RESULT 6				
ABG32185				
ID	ABG32185	standard; protein; 380 AA.		
XX	ABG32185;			
AC				
XX				
DT	05-NOV-2002	(first entry)		
XX				
DE	HCV protease NS2/3 truncation mutant 827-1206.			
XX				
KW	HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;			
KM	chronic liver disease; cirrhosis; end-stage liver disease; virus;ide;			
KM	hepatotropic; antiinflammatory; lauryldiethyamine oxide; LDO;			
KM	chaotropic agent; mutant; mutein.			
XX				
OS	Hepatitis C virus.			
OS	Synthetic.			
XX				
PN	MO200248375-A2.			
XX				
PD	20-JUN-2002.			
PF	13-DEC-2001; 2001WO-CA001796.			
XX				
PR	15-DEC-2000; 2000US-0256031P.			
XX				
PA	(BOE) BOEHRINGER INGELHEIM CANADA LTD.			
XX				
PI	Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;			

DR	XX	WP1; 2002-599511/64.	
XX	PT	Novel polypeptide for screening inhibitors of non-structural proteases	
PT	XX	useful as therapeutic agents against hepatitis C virus, comprises full	
XX	XX	length non-structural protease, or its truncation.	
PS	XX	Claim 41; Page 60-61; 67pp; English.	
CC	CC	The invention relates to an isolated polypeptide consisting of a full-	
CC	CC	length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred	
CC	CC	to also as NS2/3 (810-1206))' or its truncation, having as its N-terminal	
CC	CC	residue amino acid 810 to 906, or having a minimal amino acid sequence	
CC	CC	from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length	
CC	CC	NS2/3 protease. Also included are (1) a composition (C) comprising an	
CC	CC	isolated HCV NS2/3 protease selected from full length NS2/3 protease, or	
CC	CC	its truncation or a mutated sequence, where the protease is in a solution	
CC	CC	comprising a sufficient concentration of lauryldimethylamine oxide (LDAO)	
CC	CC	to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide	
CC	CC	appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3	
CC	CC	protease, involving isolating the protease in the presence of a	
CC	CC	chaotropic agent, refolding the isolated protease by contacting it with a	
CC	CC	reducing agent, and LDAO in the presence of reduced concentration of the	
CC	CC	chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3	
CC	CC	protease, involving diluting refolded inactive NS2/3 protease in a medium	
CC	CC	containing an activation detergent to induce auto-cleavage of the NS2/3	
CC	CC	protease; (5) measuring (M3) the auto-cleavage activity of NS2/3	
CC	CC	protease, involving incubating the active NS2/3 protease produced by M2	
CC	CC	for sufficient time to induce auto-cleavage of NS2/3 protease and produce	
CC	CC	cleavage products or their fragments, and measuring the presence or	
CC	CC	absence of uncleaved NS2/3 protease, cleavage products or their fragments	
CC	CC	; and (6) screening a potential inhibitor of auto-cleavage activity of an	
CC	CC	active NS2/3 protease, involving carrying out M3 in the presence of, or	
CC	CC	absence of the potential inhibitor, comparing the amount of uncleaved	
CC	CC	NS2/3 protease, cleavage products or their fragments. The protease is	
CC	CC	useful for detailed biochemical characterization of the enzymes and in	
CC	CC	the development of in vitro assays for screening novel inhibitors of	
CC	CC	NS2/3 protease which are useful as therapeutic agents against HCV	
CC	CC	infection (which causes chronic liver disease, cirrhosis and end-stage	
CC	CC	liver disease. M1 is useful for high level production of protease. The	
CC	CC	present sequence represents the NS2/3 truncation mutant 827-1206	
SQ	SX	(numbered relative to the full length NS2/3 protein)	
SQ	SX	Sequence 380 AA;	
Query Match	89.7%; Score 1589; DB 5; Length 380;		
Best Local Similarity	100.0%; Pred. No. 3.8e-146;		
Matches 303; Conservative	0; Mismatches 0; Indels 0; Gaps 0		
Df	Qy	16 AGITKVPFVFAOGLIRACMLVRKAAGHYIQMAFMKLAALTGTYYVTHLPLOWMALNG 75	
Df	Df	78 AGITKVPFVFAOGLIRACMLVRKAAGHYIQMAFMKLAALTGTYYVTHLPLOWMALNG 137	
Qy	76	LRLDAVAEPIVFESMEVKIITWGDYTAACGGIIISGLPVSARGRREILGPADNEEGCGM 135	
Df	138	LRLDAVAEPIVFESMEVKIITWGDYTAACGGIIISGLPVSARGRREILGPADNEEGCGM 197	
Qy	136	RLLAEITAYSOQTRGLICCIITSLTGRDXNOVEGVAVSTATSGFLATCNVCWCYTYPH 195	
Df	198	RLLAEITAYSOQTRGLICCIITSLTGRDXNOVEGVAVSTATSGFLATCNVCWCYTYPH 257	
Qy	156	GAGSCTLTAGPGPILOTMTNVDOLVGMOAPFGASMPCTCGSSDLYLVTRHADVIPR 255	
Df	258	GAGSCTLTAGPGPILOTMTNVDOLVGMOAPFGASMPCTCGSSDLYLVTRHADVIPR 317	
Qy	256	RKGDSRGSLISRPFVSYLKSGSGGLCPSGHAVGIFFAAVCTRGVAKADVFIPIVESNET 315	
Df	318	RKGDSRGSLISRPFVSYLKSGSGGLCPSGHAVGIFFAAVCTRGVAKADVFIPIVESMET 377	
Qy	316	TMR 318	
Df	378	TMR 380	

RESULT 7  
 ABG32184  
 ID ABG32184 standard; protein; 393 AA.  
 AC ABG32184;  
 XX  
 XX  
 DT 05-NOV-2002 (first entry)  
 DE HCV protease NS2/3 truncation mutant 815-1206.  
 XX  
 XX HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocid;  
 XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 OS  
 OS Hepatitis C virus.  
 OS Synthetic.  
 PN WO200248375-A2.  
 XX  
 XX 20-JUN-2002.  
 XX  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 XX Thibeault D, Lemaire D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 XX Claim 41; Page 59-60; 67pp; English.

SO Sequence 393 AA;  
 Query Match 89.7%; Score 1589; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 4e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFPAAGLIRACMLVRKAAAGHYVQMAFVKALALGTGYDHLTFLODPAHAG 75  
 DB 91 AGITKVPYFPAAGLIRACMLVRKAAAGHYVQMAFVKALALGTGYDHLTFLODPAHAG 150  
 QY 76 LRDLAAYVEPVIFSDMEVKIITWGAADTAACDIIISGLPVSAARGREIILGPADNFEQGM 135  
 DB 151 LRDLAAYVEPVIFSDMEVKIITWGAADTAACDIIISGLPVSAARGREIILGPADNFEQGM 210  
 QY 136 RLAPITTAASQOTRGLGCIITSLGRDNQVGEVQVSTATQSFATCNAVGVCTVPH 195  
 DB 211 RLAPITTAASQOTRGLGCIITSLGRDNQVGEVQVSTATQSFATCNAVGVCTVPH 270  
 QY 196 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGASMTPTCGSSDLYLVRHADVIPVR 255  
 DB 271 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGASMTPTCGSSDLYLVRHADVIPVR 330  
 QY 256 RRQDSRGSLLSRPPVSYLKSSGGLPCSGHAGVIFRAVCTRGVAAVDFIPVESNET 315  
 DB 331 RRQDSRGSLLSRPPVSYLKSSGGLPCSGHAGVIFRAVCTRGVAAVDFIPVESNET 390  
 QY 316 TWR 318  
 DB 391 TWR 393  
 RESULT 8  
 ABG32191  
 ID ABG32191 standard; protein; 303 AA.  
 AC ABG32191;  
 XX  
 XX 05-NOV-2002 (first entry)  
 XX  
 XX HCV protease NS2/3 truncation 904-1206/Cys993A1A.  
 XX  
 XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocid;  
 XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 OS  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 90 /note="Wild-type Cys substituted by Ala"  
 XX  
 XX WO200248375-A2.  
 XX  
 XX 20-JUN-2002.  
 XX  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 XX Thibeault D, Lemaire D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 XX Disclosure; Page 65-66; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant Cys932Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

Sequence 303 AA;

Query Match 89.2%; Score 1580; DB 5; Length 303;  
 Best Local Similarity 99.7%; Pred. No. 2,1e-145;  
 Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

16 AGITVPEFVFAOGLIRACMLVRKAAGHYVQMAEMKLAULTGVYDHLTPLODMARAG 75  
 1 AGITVPEFVFAOGLIRACMLVRKAAGHYVQMAEMKLAULTGVYDHLTPLODMARAG 60

76 LRDLAANPEVTFSEMEVKITWGDYTAAGDITISGLVPSARRGRHILLGPDNFEQGW 135  
 61 LRDLAANPEVTFSEMEVKITWGDYTAAGDITISGLVPSARRGRHILLGPDNFEQGW 120

136 RLALATITVYSQOTRGLGCIITSLGRBDKNQVEGVQVSTATQSFILATCNVGCWTFH 195  
 121 RLALATITVYSQOTRGLGCIITSLGRBDKNQVEGVQVSTATQSFILATCNVGCWTFH 180

196 GAGSKTLAPGPKPITQMTNTVDQDLYGVQAPPGARSMPTCTCGSSDLYLVTRHADYIVR 255  
 181 GAGSKTLAPGPKPITQMTNTVDQDLYGVQAPPGARSMPTCTCGSSDLYLVTRHADYIVR 240

256 RRGDSRGSLSPRPVSYLKSGSGGPLLCPSGHVAIFFAAVTREGVAVAVPIPVESMET 315  
 241 RRGDSRGSLSPRPVSYLKSGSGGPLLCPSGHVAIFFAAVTREGVAVAVPIPVESMET 300

316 TMR 318  
 301 TMR 303

RESULT 9  
 ABG32189  
 ID ABG32189 standard; protein; 303 AA.  
 AC ABG32189;  
 XX  
 XX  
 DT 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation 904-1206/HIS952Ala.  
 DE HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
 KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KW chaotropic agent; mutant; muten.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT Misc-difference 49 /note="Wild-type His substituted by Ala"  
 XX  
 XX MO200248375-A2.  
 XX  
 XX 20-JUN-2002.  
 XX  
 XX 13-DEC-2001; 2001MO-CA001796.  
 XX  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX  
 XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 PA  
 PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX  
 XX WPI; 2002-599511/64.  
 DR  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PT  
 XX  
 PS Example 7, Fig 8; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant HIS952Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

Query Match 89.2%; Score 1579; DB 5; Length 303;  
 Sequence 303 AA;

Best Local Similarity 99.7%; Pred No. 2.6e-145; Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGIIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75
DB 1 AGITKVPYFVRAQGIIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60
QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPDNFEQGM 135
DB 61 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPDNFEQGM 120
QY 136 RLAPITAVSQOTRGLGCIITSLTRGRDNQVEGVQVSTATOSFLATCVNGVMTVPH 195
DB 121 RLAPITAVSQOTRGLGCIITSLTRGRDNQVEGVQVSTATOSFLATCVNGVMTVPH 180
QY 196 GAGSKTLGPKGPITQMTNTNDODLVGQAPPGARSMTPTCGSSDLYLVTRHADVPVR 255
DB 181 GAGSKTLGPKGPITQMTNTNDODLVGQAPPGARSMTPTCGSSDLYLVTRHADVPVR 240
QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPSSGHAIVGIFPAAVCTRGVAKAVDFIVESEMET 315
DB 241 RRGDSRGSILSPRPVSYLKSSGGPILCPSSGHAIVGIFPAAVCTRGVAKAVDFIVESEMET 300
QY 316 TMR 318
DB 301 TMR 303

```

## RESULT 10

ABG32190  
ID ABG32190 standard; protein, 301 AA.

AC ABG32190;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation 904-1206/deltaLeu1026-Ala1027.

KW HCV, enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 KW chronic liver disease; cirrhosis; end-stage liver disease; viruslike;  
 KW hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KW chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

FX Key Location/Qualifiers

FT Misc-difference 122..123

FT /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"

FN WO200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CAC01796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

DR WPI; 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

PS Example 7: Page 64-65; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal

CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation 904-1206 mutant  
 CC deltaLeu1026-Ala1027 (numbered relative to the full length NS2/3 protein)  
 CC a mutant devoid of autocatalytic activity  
 CC SX

## Sequence 301 AA;

Query Match 88.7%; Score 1570; DB 5; Length 301;

Best Local Similarity 99.3%; Pred No. 1.9e-144; Matches 301; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

```

QY 16 AGITKVPYFVRAQGIIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75
DB 1 AGITKVPYFVRAQGIIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60
QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPDNFEQGM 135
DB 61 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPDNFEQGM 120
QY 136 RLAPITAVSQOTRGLGCIITSLTRGRDNQVEGVQVSTATOSFLATCVNGVMTVPH 195
DB 121 RLAPITAVSQOTRGLGCIITSLTRGRDNQVEGVQVSTATOSFLATCVNGVMTVPH 178
QY 196 GAGSKTLGPKGPITQMTNTNDODLVGQAPPGARSMTPTCGSSDLYLVTRHADVPVR 255
DB 179 GAGSKTLGPKGPITQMTNTNDODLVGQAPPGARSMTPTCGSSDLYLVTRHADVPVR 238
QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPSSGHAIVGIFPAAVCTRGVAKAVDFIVESEMET 315
DB 239 RRGDSRGSILSPRPVSYLKSSGGPILCPSSGHAIVGIFPAAVCTRGVAKAVDFIVESEMET 298
QY 316 TMR 318
DB 299 TMR 301

```

## RESULT 11

ABG32188  
ID ABG32188 standard; protein, 292 AA.

AC ABG32188;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 915-1206.

DB

KW HCV; enteric protease; NS2/3 (915-1206); hepatitis C virus infection;  
KM chronic liver disease; cirrhosis; end-stage liver disease; vintedide;  
KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutlein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
PN MO200248375-A2.  
PD 20-JUN-2002.  
PF 13-DEC-2001; 2001MO-CAN01796.  
PR 15-DEC-2000; 2000US-0256031P.  
PX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
PY Thibault D, Lamarre D, Maurice R, Pilote I., Pause A;  
PI WPI; 2002-599511/64.  
PP Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
XX  
XX Claim 41; Page 63; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS2/3) protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as AEG33198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 915-1206 (numbered relative to the full length NS2/3 protein)

Sequence 292 AA:  
SQ

Query Match 86.5%; Score 1532; DB 5; Length 292;  
Best Local Similarity 100.0%; Pred. No. 9.5e-11;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0

27 AAGIIRKCMVRRAGSHYQMAFMKLALNSTYYDHLTPLOMAAAGRDIAVAPEV 86  
Db 1 AAGIIRKCMVRRAGSHYQMAFMKLALNSTYYDHLTPLOMAAAGRLDAVAPEV 60

87 IFSDMEKXITTWGADTAAAGDIISGLPVASARRGBEILGPADFEGQGWLAPITAYSQ 146

D6		61	IFSDMEVKLIITMGADTAAAGDIIISGLPVSARSGREILLGPADNPEGGMFLAIPITAYSQ	120
OY		147	QTRGILLCITTSITGRDKRQVGEFQVAVSTATQSFIATCNVNGWMTVFHQAASKTLAEPK	206
D6		121	QTREGLLGCITITSITGDKRQVGEFQVAVSTATQSFIATCNVNGWMTVFHQAASKTLAEPK	180
OY		207	GPIITQWTNVVDODLVGWQAPGARASMTPECTCGSSDLVLVTSHADVIVERRKDSRGSLLS	266
D6		181	GIPTQWTNVTVDDIVGWQAPGARASMTPECTCGSSDLVLVTSHADVIVERRKDSRGSLLS	240
OY		267	PREVSVLKSSGGGPLLCPGSGNAVGIFRAAVCTRGVAKAVDPIPVESMETTMK	318
D6		241	PREVSVLKSSGGGPLLCPGSGNAVGIFRAAVCTRGVAKAVDPIPVESMETTMK	292
 RESULT 12 ABG30601 ID ABG30601 standard; protein; 2201 AA.				
XX			ABG30601;	
AC				
XX				
DT		21-OCT-2002	(first entry)	
XX				
DE		Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.		
XX				
KW		Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.		
OS		Hepatitis C virus. Synthetic.		
XX				
FH		Key	Location/Qualifiers	
FT		Misc-difference 882	/label= Arg, Lys	
FT		Misc-difference 2183		
FT			/note= "Wild type Met substituted by Thr"	
FN		WO200252015-A2.		
PD		04-JUL-2002.		
PF		20-DEC-2001; 2001MO-CA001843.		
PR		22-DEC-2000; 2000US-0257857P.		
XX				
PA		(BOEH ) BOEHRINGER INGELHEIM CANADA LTD.		
PI		Kukolj G, Pause A;		
XX				
XX		WPI; 2002-575382/61.		
DR				
PT		New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.		
PS		Claim 3; Page; 140pp; English.		
XX				
CC		The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self- replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self- replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in construction with mutations in the HCV non-structural region, such as the G12042/C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon ARGX2 and contains the viral protease NS2/3, protease C virus replicon ARGX2 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:		

CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

XX Sequence 2201 AA:

Query Match 86.4%; Score 1531; DB 5; Length 2201;  
 Best Local Similarity 94.7%; Pred. No. 2.1e-139;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```

QY 16 AGITKVPFVFAAGLIRACMLVRKAGAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 75
DB 95 AGITKVPFVFAAGLIRACMLVRKAGAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 154
QY 76 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135
DB 155 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 214
QY 136 RLAPITAYVSOQTRGLLGCITTSITLGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 195
DB 215 RLAPITAYVSOQTRGLLGCITTSITLGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 274
QY 196 GAGSKTLAGPKGPIITQWNTYNDQDLYVGWQAPPGARSMTPCTCGSSDLYLTRHADVI 255
DB 275 GAGSKTLAGPKGPIITQWNTYNDQDLYVGWQAPPGARSMTPCTCGSSDLYLTRHADVI 334
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKADVIVVESMET 315
DB 335 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKADVIVVESMET 394
QY 316 TWRT 319
DB 395 TWR 398

```

#### RESULT 13

ABG30591 standard; protein; 2201 AA.

ABG30591;

21-OCT-2002 (first entry)

Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.

Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

Hepatitis C virus.

Synthetic.

Location/Qualifiers

Misc-difference 751 /note= "Wild type Ser substituted by Gly"

Misc-difference 882 /label= Arg, Lys

WO200252015-A2.

04-JUL-2002.

20-DEC-2001; 2001WO-CA001843.

22-DEC-2000; 2000US-0257857P.

(BOEH) BOEHRINGER INGELHEIM CANADA LTD.

Kukolj G, Pause A;

WPI; 2002-575382/61.

New self-replicating RNA molecules from Hepatitis C virus (HCV), which

possess enhanced transduction or replication efficiency, useful for

evaluating potential inhibitors of HCV replication.

XX Claim 3; Page; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G12042C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC virus replicon Apck12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

XX Sequence 2201 AA:

Query Match 86.4%; Score 1531; DB 5; Length 2201;  
 Best Local Similarity 94.7%; Pred. No. 2.1e-139;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```

QY 16 AGITKVPFVFAAGLIRACMLVRKAGAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 75
DB 95 AGITKVPFVFAAGLIRACMLVRKAGAGHYQMAFMKLAALTGYVVDHLTPLDMAHAG 154
QY 76 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135
DB 155 LRDIAVAEVPVIFSDMEVKIITMGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 214
QY 136 RLAPITAYVSOQTRGLLGCITTSITLGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 195
DB 215 RLAPITAYVSOQTRGLLGCITTSITLGRDNQVEGEVQVSTATQSFATCNGVCWTVYH 274
QY 196 GAGSKTLAGPKGPIITQWNTYNDQDLYVGWQAPPGARSMTPCTCGSSDLYLTRHADVI 255
DB 275 GAGSKTLAGPKGPIITQWNTYNDQDLYVGWQAPPGARSMTPCTCGSSDLYLTRHADVI 334
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKADVIVVESMET 315
DB 335 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKADVIVVESMET 394
QY 316 TWRT 319
DB 395 TWR 398

```

#### RESULT 14

ABG30600 standard; protein; 2201 AA.

ABG30600;

21-OCT-2002 (first entry)

Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

Hepatitis C virus.

Synthetic.

Location/Qualifiers

Misc-difference 882 /label= Arg, Lys

ABG30600 standard; protein; 2201 AA.

FT Misc-difference 1357  
 FT /note= "Wild type Pro substituted by Leu"  
 XX  
 FN W0200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page: 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G12042/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not represent the specific information but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX  
 XX Sequence 2201 AA;  
 SQ  
 Query Match 86.4%; Score 1531; DB 5; Length 2201;  
 Best Local Similarity 94.7%; Pred. No. 2,1e-139;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

RESULT 15  
 ID ABG30581 standard; protein; 2201 AA.  
 XX  
 AC ABG30581;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.  
 XX  
 KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PN W0200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 DR N-PSDB; ABK68573.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Disclosure; Page 49-58; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G12042/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apk12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 CC  
 XX  
 XX Sequence 2201 AA;  
 SQ  
 Query Match 86.4%; Score 1531; DB 5; Length 2201;  
 Best Local Similarity 94.7%; Pred. No. 2,1e-139;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY	256	RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET	315
Db	335	RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET	394
QY	316	TMRT	319
Db	395	TMRG	398

Search completed: May 6, 2004, 09:30:43  
Job time : 46.463 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 10.8777 Seconds

(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771

Sequence: 1 MKKKLHHHHHTSAGITK.....TTWRTSSAMRHPQGGKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1533	86.6	3010	1 A45573	genome polypotein
2	1528	86.3	3010	1 GNMVCT	genome polypotein
3	1515	85.5	3010	1 GNMVCT	genome polypotein
4	1487	84.0	3010	1 S18030	genome polypotein
5	1479	83.5	3010	1 GNMVCT	genome polypotein
6	1404	79.3	3011	1 S40770	genome polypotein
7	1399	79.0	3011	1 GNMVCT	genome polypotein
8	1386	78.3	3011	1 GNMVCT	genome polypotein
9	1235	69.7	3014	1 J56200	genome polypotein
10	1173	66.2	3033	1 J01303	genome polypotein
11	1158	65.4	3033	1 GNMVCT	genome polypotein
12	398.5	22.5	3005	1 T08841	polypotein - dour
13	343	19.4	2970	2 T08839	polypotein - marm
14	108	6.1	353	2 G87392	DNA-directed DNA P
15	101	5.7	600	2 B46462	conserved hypochet
16	98.5	5.6	470	2 JC4058	tetracycline 6-hyd
17	97.5	5.5	1085	2 T03531	coBN protein homol
18	95.5	5.4	706	2 S33761	transferin precur
19	94.5	5.3	716	2 G83612	hypothetical prote
20	93.5	5.3	7463	2 T36248	CDA peptidase synthe
21	93	5.3	660	2 VHMW2	structural protein
22	91	5.1	904	2 A84212	hypothetical prote
23	90.5	5.1	868	2 H81775	acetylcholinesteras
24	90	5.1	659	2 B44212	structural protein
25	90	5.1	2796	2 JC4743	fatty-acid synthas
26	89.5	5.1	267	2 B83602	conserved hypochet
27	88	5.0	3414	1 GNMVCT	genome polypotein
28	87	4.9	1057	2 T18171	hngl protein - hum
29	87	4.9	3412	1 GNMVCT	genome polypotein

30	86.5	4.9	470	1 NMIVW8	exo-alpha-sialidas
31	86.5	4.9	990	2 S67499	glutamate synthase
32	85.5	4.8	299	2 AH3447	cytochrome-c oxida
33	85.5	4.8	348	2 H70549	probable pnh prot
34	85	4.8	470	1 NMIV9	exo-alpha-sialidas
35	85	4.8	707	2 D84154	cadum-transporci
36	84.5	4.8	347	2 S44167	malate dehydrogen
37	84.5	4.8	5627	2 C83339	hypothetical prote
38	84	4.7	223	2 T35594	hypothetical prote
39	84	4.7	3069	2 H70656	fatty-acid synthas
40	83.5	4.7	315	2 AG2361	hypothetical prote
41	83.5	4.7	538	2 S22409	D-alanyl-D-alanine
42	83.5	4.7	1399	2 G83112	DNA-directed RNA P
43	83	4.7	398	2 B71284	probable periplasm
44	83	4.7	2103	2 G86925	probable polyketid
45	83	4.7	4735	2 T17463	flamycin polyketi

#### ALIGNMENTS

##### RESULT 1

A45573  
genome polypotein - hepatitis C virus (strain J7)  
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C/Accession: A45573  
R:tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, I  
Virus Res. 23, 39-53, 1992  
A/Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s  
A/Reference number: A45573; PMID:92295714; PMID:1318627  
A/Accession: A45573  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-3010 <TRAN>  
A/Cross-references: GB:D11168; GB:D011171; NID:9221612; PID:BA001943.1; PID:9221613  
A/Experimental source: HCV-JT  
A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIPI:106207)  
C:Keywords: ATP; glycoprotein; hydroxase; nucleotide binding; P-loop; polypotein; serin  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPN>  
F:192-389/Product: major envelope protein B #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>  
F:1120-1237/Region: nucleotide-binding motif A (P-loop)  
F:112-1317/Region: nucleotide-binding motif B  
F:116-1319/Region: DEKH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1663-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 86.6%; Score 1533; DB 1; Length 3010;  
Best Local Similarity 95.1%; Pred. No. 2.7e-121;  
Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY	16	AGITKPYEVRAGGLIRACMLVKAAGHYVQMAFKLALITGYTYDHLTPLODAHAG	75
DB	904	AATLAPYFVRAGGLIRACMLVKAAGHYVQMAFKLALITGYTYDHLTPLODAHAG	963
QY	76	LRLAVAVEPVIFSDMEVKITWGAATAACGDIISLPSARGRHILIGPDNPEGQGM	135
DB	964	LRLAVAVEPVIFSDMEVKITWGAATAACGDIISLPSARGRHILIGPDNPEGQGM	1023
QY	136	RLAAPTAYSOQRLAGCITSLTRDKNOVEGEVQVSTATQSLATCNGVGVCTVH	195
DB	1024	RLAAPTAYSOQRLAGCITSLTRDKNOVEGEVQVSTATQSLATCNGVGVCTVH	1083
QY	196	GAGSKTLAPKGBITQMTYNNVDOLVWQAPPGARSMTCTCGSSDLVYVTRHADVIVR	255

Db 1084 GAGSXTLAGPKGITOMYTNVDQDLVGMHAPPGARSILPTCGSSDLYLTVRHADYIPVR 1143  
 |||  
 QY 256 RRGDSRGSILSPRPVSYLKSGSGGLPCPSGHAVGIFRAAVCTRGVAKADVPIVESMET 315  
 |||  
 Db 1144 RRGDSRGSILSPRPVSYLKSGSGGLPCPSGHAVGIFRAAVCTRGVAKADVPIVESMET 1203  
 |||  
 QY 316 TMRT 319  
 |||  
 Db 1204 TMRS 1207

## RESULT 2

GNMVTW  
 genome polyprotein - hepatitis C virus (strain J)  
 N:contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
 C/Accession: A39253; PM0086  
 R/Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
 A/Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
 A/Reference number: A39253; MUID:91088550; PMID:2175903  
 A/Accession: A39253  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <KAT>  
 A/Cross-references: GB:D90208; NID:G221610; PIDN:BA14233.1; PID:G221611  
 R/Kato, N.; Ohkoshi, S.; Shimotohno, K.  
 Proc. Jpn. Acad. 65B, 219-223, 1989  
 A/Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari  
 A/Reference number: PS0085  
 A/Accession: PS0086  
 A/Molecule type: genomic RNA  
 A/Residues: 2650-2707 <KAT>  
 A/Experimental source: Japanese isolate  
 C/Comment: The cleavage sites of this polyprotein have not been determined.  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F/2-115/Product: capsid protein C #status predicted <CPC>  
 F/119-191/Product: envelope protein M #status predicted <EPM>  
 F/136-1319/Product: major envelope protein E #status predicted <MEE>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1319/Region: DEXH motif  
 F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,223,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 86.3%; Score 1528; DB 1; Length 3010;  
 Best Local Similarity 94.1%; Pred. No. 7,1e-121;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVDHLTPLODMAHAG 75  
 |||  
 Db 904 AGITRVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVDHLTPLODMAHAG 963  
 |||  
 QY 76 LRDIAVAEVPYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
 |||  
 Db 964 LRDIAVAEVPYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 1023  
 |||  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVGEVQVASTATOSFLATCNVGCMTVEH 195  
 |||  
 Db 1024 RLAPITAYSOOTRGLGCIITSLTGRDNQVGEVQVASTATOSFLATCNVGCMTVEH 1083  
 |||  
 QY 196 GAGSKTLAAGPKGITOMYTNVDQDLVGMHAPPGARSMTPTCGSSDLYLTVRHADYIPVR 255  
 |||  
 Db 1084 GAGSKTLAAGPKGITOMYTNVDQDLVGMHAPPGARSMTPTCGSSDLYLTVRHADYIPVR 1143  
 |||  
 QY 256 RRGDSRGSILSPRPVSYLKSGSGGLPCPSGHAVGIFRAAVCTRGVAKADVPIVESMET 315

Db 1144 RRGDSRGSILSPRPVSYLKSGSGGLPCPSGHAVGIFRAAVCTRGVAKADVPIVESMET 1203  
 |||  
 QY 316 TMRT 319  
 |||  
 Db 1204 TMRS 1207

## RESULT 3

GNMVTW  
 genome polyprotein - hepatitis C virus (strain Taiwan)  
 N:contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 A/Note: host Homo sapiens (man)  
 C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C/Accession: A40244  
 R/Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
 Virology 189, 102-113, 1992  
 A/Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A/Reference number: A40244; MUID:92230206; PMID:1314449  
 A/Accession: A40244  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <CHE>  
 A/Cross-references: GB:M8754  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <MEE>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1319/Region: DEXH motif  
 F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,223,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 85.5%; Score 1515; DB 1; Length 3010;  
 Best Local Similarity 92.8%; Pred. No. 9e-120;  
 Matches 282; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVDHLTPLODMAHAG 75  
 |||  
 Db 904 AGITRVPYFVRAOGILRACMLVRKAAGHYVOMAFKLAALTGTVYVDHLTPLODMAHAG 963  
 |||  
 QY 76 LRDIAVAEVPYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
 |||  
 Db 964 LRDIAVAEVPYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 1023  
 |||  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVGEVQVASTATOSFLATCNVGCMTVEH 195  
 |||  
 Db 1024 RLAPITAYSOOTRGLGCIITSLTGRDNQVGEVQVASTATOSFLATCNVGCMTVEH 1083  
 |||  
 QY 196 GAGSKTLAAGPKGITOMYTNVDQDLVGMHAPPGARSMTPTCGSSDLYLTVRHADYIPVR 255  
 |||  
 Db 1084 GAGSKTLAAGPKGITOMYTNVDQDLVGMHAPPGARSMTPTCGSSDLYLTVRHADYIPVR 1143  
 |||  
 QY 256 RRGDSRGSILSPRPVSYLKSGSGGLPCPSGHAVGIFRAAVCTRGVAKADVPIVESMET 315  
 |||  
 Db 1144 RRGDSRGSILSPRPVSYLKSGSGGLPCPSGHAVGIFRAAVCTRGVAKADVPIVESMET 1203  
 |||  
 QY 316 TMRT 319  
 |||  
 Db 1204 TMRS 1207

RESULT 4  
 S18030  
 genome polyprotein - hepatitis C virus (isolate JX1)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Variety: isolate JKI  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
 C:Accession: S18030; S33570; A48332; S18029  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie  
 A:Reference number: S18028  
 A:Accession: S18030  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <HON>  
 A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G959479  
 A:Experimental source: isolate JKI from an individual  
 R:Honda, M.; Kaneko, S.; Ueno, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 148, 163-169, 1993  
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A:Reference number: A48332; MUID:93119270; PMID:8380322  
 A:Accession: S33570  
 A:Molecule type: genomic RNA  
 A:Residues: 1-547, 'T', 548-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
 A:Cross-references: EMBL:X61591  
 A:Note: this sequence is inconsistent with the nucleotide translation  
 A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320  
 as Trp, and TTC for residue 771 as Ser  
 A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:P:121748)  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEKH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (as

Query Match 84.0%; Score 1487; DB 1; Length 3010;  
 Best Local Similarity 92.1%; Pred. No. 2.1e-117;  
 Matches 280; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 Db 904 AGITRVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 QY 76 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 135  
 Db 964 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 1023  
 QY 136 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVYH 195  
 Db 1024 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVYH 1083  
 QY 196 GAGSKTLAPKPGITOMNTYNDVLGMOAPPGARSMTPCTCGSSDLYLVTRHADVPVR 255  
 Db 1084 GAGSKTLAPKPGITOMNTYNDVLGMOAPPGARSMTPCTCGSSDLYLVTRHADVPVR 1143  
 QY 256 RRGDSRGSLSRPVSYLKGSGGFLPCPSGHAVGIRAAVCTRGVAKAVDFIPVESMET 315  
 Db 1144 RRGDSRGSLSRPVSYLKGSGGFLPCPSGHAVGIRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMR 319  
 Db 1204 TMR 1207

RESULT 5

GNM/TC  
 genome polyprotein - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
 C:Accession: A38465  
 R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J  
 U. Virol. 65, 1105-1113, 1991  
 A:Title: Structure and organization of the hepatitis C virus genome isolated from human  
 A:Reference number: A38465; MUID:91140698; PMID:1847440  
 A:Accession: A38465  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <TA>  
 A:Cross-references: EMBL:X68335; NID:G329770; PIDN:AAA72945.1; PID:G329771  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEKH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,425,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22,

Query Match 83.5%; Score 1479; DB 1; Length 3010;  
 Best Local Similarity 91.8%; Pred. No. 1e-116;  
 Matches 279; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 Db 904 AGITRVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 QY 76 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 135  
 Db 964 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNFEQGW 1023  
 QY 136 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVYH 195  
 Db 1024 RLAPITRAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCNGVCWTVYH 1083  
 QY 196 GAGSKTLAPKPGITOMNTYNDVLGMOAPPGARSMTPCTCGSSDLYLVTRHADVPVR 255  
 Db 1084 GAGSKTLAPKPGITOMNTYNDVLGMOAPPGARSMTPCTCGSSDLYLVTRHADVPVR 1143  
 QY 256 RRGDSRGSLSRPVSYLKGSGGFLPCPSGHAVGIRAAVCTRGVAKAVDFIPVESMET 315  
 Db 1144 RRGDSRGSLSRPVSYLKGSGGFLPCPSGHAVGIRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMR 319  
 Db 1204 TMR 1207

RESULT 6

S40770  
 genome polyprotein - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: S40770; PCL285  
 R:Okamoto, H.  
 submitted to the EMBL Data Library, March 1992  
 A:Reference number: S40770  
 A:Accession: S40770  
 A:Molecule type: genomic RNA

A/Residues: 1-3011 <OKA>  
 A/Cross-references: EMBL:U00749; NID:G221586; PIDN:BA01582.1; PID:G221587  
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Ugn. J. Exp. Med. 60, 167-177, 1990  
 A/Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A/Reference number: PC1284; MUID:91013116; PMID:2170712  
 A/Accession: PC1285  
 A/Molecule type: genomic RNA  
 A/Residues: 1-513 <OK2>  
 A/Cross-references: GB:PD0831; NID:G221511; PIDN:BA00705.1; PID:G221512  
 A/Experimental source: isolate HC-01  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; hydrolyase; nucleotide binding; P-loop; polyprotein; serin  
 F/2-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <MEP>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 Query Match 79.3%; Score 1404; DB 1; Length 3011;  
 Best Local Similarity 84.5%; Pred. No. 2.4e-110;  
 Matches 257; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITKVPFVVRVNOGILRACMLVRKAGHYQVAFMCLALTGVYVDHLTPLODMAHG 75  
 DB 904 ASLKVPEFVRVQGLRICALARKVQGHVQVMIKLGALTGVYVNHLPDMAHNG 963  
 QY 76 LRDLAVAEVPVIFSDMEVKIITWGADTAACDIIISGLPVSARKREIILGPADNFBGQW 135  
 DB 964 LRDLAVAEVPVIFSDMEVKIITWGADTAACDIIISGLPVSARKREIILGPADNFBGQW 1023  
 QY 136 RLAPITAVSOQTRGLGCIITSLTGPRKNOVEGVQVVSATOSFLATCVNGVCTVFEH 155  
 DB 1024 RLAPITAVSOQTRGLGCIITSLTGPRKNOVEGVQVVSATOSFLATCVNGVCTVFEH 1083  
 QY 196 GAGSKTLGPKPIQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 255  
 DB 1084 GAGRTITLSPKGPVYQMTNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLYKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSILSPRPVSYLYKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 7  
 GNMVCH  
 genome polyprotein - hepatitis C virus (strain HCV-1)  
 N/contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/species: hepatitis C virus  
 C/date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 19-Jan-2001  
 A/Accession: A39166; PQ0403; PQ0404  
 R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A/Title: Genetic organization and diversity of the hepatitis C virus.  
 A/Reference number: A39166; MUID:91172826; PMID:1848704  
 A/Accession: A39166  
 A/Molecule type: mRNA  
 A/Residues: 1-3011 <COC>  
 A/Cross-references: GB:M62321; NID:G329873; PIDN:AA45676.1; PID:G329874  
 R/Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992

A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A/Reference number: PQ0393; MUID:92268871; PMID:1316939  
 A/Accession: PQ0403  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CHA>  
 A/Cross-references: DBJ:U01028  
 A/Experimental source: isolates E-b16  
 A/Accession: PQ0404  
 A/Status: preliminary  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CH2>  
 A/Experimental source: isolates E-b17  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolyase; nonstructural  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: envelope protein M #status predicted <EPM>  
 F/192-389/Product: major envelope protein E #status predicted <MEP>  
 F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F/1312-1317/Region: nucleotide-binding motif B  
 F/1316-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,222  
 Query Match 79.0%; Score 1399; DB 1; Length 3011;  
 Best Local Similarity 84.2%; Pred. No. 6.3e-110;  
 Matches 256; Conservative 25; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITKVPFVVRVNOGILRACMLVRKAGHYQVAFMCLALTGVYVDHLTPLODMAHG 75  
 DB 904 ASLKVPEFVRVQGLRICALARKVQGHVQVMIKLGALTGVYVNHLPDMAHNG 963  
 QY 76 LRDLAVAEVPVIFSDMEVKIITWGADTAACDIIISGLPVSARKREIILGPADNFBGQW 135  
 DB 964 LRDLAVAEVPVIFSDMEVKIITWGADTAACDIIISGLPVSARKREIILGPADNFBGQW 1023  
 QY 136 RLAPITAVSOQTRGLGCIITSLTGPRKNOVEGVQVVSATOSFLATCVNGVCTVFEH 195  
 DB 1024 RLAPITAVSOQTRGLGCIITSLTGPRKNOVEGVQVVSATOSFLATCVNGVCTVFEH 1083  
 QY 196 GAGSKTLGPKPIQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 255  
 DB 1084 GAGRTITLSPKGPVYQMTNTVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLYKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSILSPRPVSYLYKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 8  
 GNMVCH  
 genome polyprotein - hepatitis C virus (strain H)  
 N/contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/species: hepatitis C virus  
 A/Note: host Homo sapiens (man)  
 C/date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 19-Jan-2001  
 A/Accession: A36814; A41546  
 R/Inchausti, G.; Zepede, S.; Lee, D.H.; Sugtani, M.; Nasoff, M.; Prince, A.M.  
 Submitted to GenBank, July 1992  
 A/Description: Genomic structure of the human prototype strain H of hepatitis C virus.  
 A/Reference number: A36814  
 A/Accession: A36814  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
 R:Inchausti, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
 A:Reference number: J015166; MUID:92052256; PMID:1658800  
 A:Contents: annotation  
 A:Note: neither amino acid nor nucleotide sequence is given  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein C #status predicted <CPC>  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 78.3%; Score 1386; DB 1; Length 3011;  
 Best Local Similarity 83.6%; Pred. No. 8e-109;  
 Matches 254; Conservative 27; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITVFYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 75  
 Db 904 ASLKVYFVAVQGLMLICARLKHAGHYVMAITGLTGVYVNHMLAPLDMAHAG 963

QY 76 LRDAVAVEPIFSDMEVKIITWQADTACGDIISGLPVASRRREIILGPADNFEQGM 135  
 Db 964 LRDAVAVEPIFSDMEVKIITWQADTACGDIISGLPVASRRREIILGPADNFEQGM 1023

QY 136 RLAPITVYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 135  
 Db 1024 RLAPITVYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 1083

QY 196 GAGSKTLAEPKPTITQWYVNDODLVGMOAPPGASMTPTCGSSDLYLVRHADVI 255  
 Db 1084 GAGSKTLAEPKPTITQWYVNDODLVGMOAPPGASMTPTCGSSDLYLVRHADVI 1143

QY 256 RRGDSRGLSPRPVSTYLKSGSGPILCPGSHANGIRAAVCTRGVAKADVIVESMET 315  
 Db 1144 RRGDSRGLSPRPVSTYLKSGSGPILCPGSHANGIRAAVCTRGVAKADVIVESMET 1203

QY 316 TMR 319  
 Db 1204 TMR 1207

RESULT 9  
 J03620 genome polyprotein - hepatitis C virus (isolate EUH1480)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: J03620  
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
 A:Reference number: J03620; MUID:97365693; PMID:9223423  
 A:Accession: J03620  
 A:Molecule type: mRNA  
 A:Residues: 1-3014 <CHA>  
 A:Cross-references: GB:Y13184  
 A:Experimental source: genotype 5a, which predominates in South Africa  
 A:Note: the translation of the nucleotide sequence is not complete in this paper  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:384-408/Region: hypervariable #status predicted  
 F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1008-1616/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
 F:1313-1318/Region: nucleotide-binding motif B  
 F:1317-1320/Region: DEXH motif  
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4a>  
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 69.7%; Score 1235; DB 1; Length 3014;  
 Best Local Similarity 72.5%; Pred. No. 5.3e-96;  
 Matches 219; Conservative 44; Mismatches 39; Indels 0; Gaps 0;

QY 18 ITXVFYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 77  
 Db 907 LTXVFYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 966

QY 78 DLVAVEPIFSDMEVKIITWQADTACGDIISGLPVASRRREIILGPADNFEQGM 137  
 Db 967 ELVAVEPIFSDMEVKIITWQADTACGDIISGLPVASRRREIILGPADNFEQGM 1026

QY 138 LAPITVYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 137  
 Db 1027 LAPITVYVFAOGLIRACMLVRKAAGHYVMAFMKLAALTGYVYDHLTPLODMAHAG 1086

QY 198 GSKTLAEPKPTITQWYVNDODLVGMOAPPGASMTPTCGSSDLYLVRHADVI 257  
 Db 1087 GSKTLAEPKPTITQWYVNDODLVGMOAPPGASMTPTCGSSDLYLVRHADVI 1146

QY 258 RRGDSRGLSPRPVSTYLKSGSGPILCPGSHANGIRAAVCTRGVAKADVIVESMET 317  
 Db 1147 RRGDSRGLSPRPVSTYLKSGSGPILCPGSHANGIRAAVCTRGVAKADVIVESMET 1206

QY 318 RT 319  
 Db 1207 RS 1208

RESULT 10  
 J03303 genome polyprotein - hepatitis C virus (isolate HC-06)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C:Accession: J03303  
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum  
 A:Reference number: J03303; MUID:92044440; PMID:1658196  
 A:Accession: J03303  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3033 <OKA>  
 A:Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651  
 A:Experimental source: isolate HC-06 from a Japanese individual  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transme  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1316-1321/Region: nucleotide-binding motif A (P-loop)  
 F:1320-1323/Region: DEXH motif  
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4a>  
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 66.2%; Score 1173; DB 1; Length 3033;  
Best Local Similarity 69.2%; Pred. No. 9,8e-91;  
Matches 209; Conservative 45; Mismatches 48; Indels 0; Gaps 0;

QY 18 ITKVFYFVAQGLIRACMLVRKAAGHYVQMAFMKLAALGTGYVDHLTPIDQMAHAGLR 77

DB 910 IIRVPEFVAHALLRMCNTVRHLAAGRYQWVLLALGRMTGYIYDHLTPMSDMAANGLR 969

QY 78 DLAAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARRREILLGANDPEQGM 137

DB 970 DLAAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARRREILLGANDPEQGM 1029

QY 138 LAPITVYSGQRTGLGCIITSLTGRDKQVGEVQVSTATQSFATCVNGVCMVTFH 197

DB 1030 LAPITVYSGQRTGLGCIITSLTGRDKQVGEVQVSTATQSFATCVNGVCMVTFH 1089

QY 198 GSKTLAAGPKGPIITQWYTNVDQVLGWAQPPGARSMTPTCGSSDLVYTRHADVI 257

DB 1090 GSKTLAAGPKGPIITQWYTNVDQVLGWAQPPGARSMTPTCGSSDLVYTRHADVI 1149

QY 258 GDSRGLSPRPVSYLKSGSGGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMET 317

DB 1150 GDSRGLSPRPVSYLKSGSGGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMET 1209

QY 318 RT 319

DB 1210 RS 1211

# RESULT 11

GNMWV8

genome polyprotein - hepatitis C virus (strain HC-J8)

N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 19-Jan-2001

C:Accession: A40250; F00397; F00359

C:Okamoto, H.; Kura, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;

Virology 188, 331-341, 1992

A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep

A:Reference number: A40250; MUID:92230232; PMID:1314459

A:Accession: A40250

A:Molecule type: genomic RNA

A:Residues: 1-3033 <OKA>

A:Cross-references: GB:DI0988; GB:DI01221; NID:g221608; PIDN:BA01761.1; PID:g221609

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A:Reference number: F00393; MUID:92268871; PMID:1316939

A:Accession: F00397

A:Molecule type: genomic RNA

A:Residues: 2678-2754 <CHA>

A:Cross-references: DDBJ:DI0134

A:Experimental source: isolate E-B-2

R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno

Biocem. Biophys. Res. Commun. 181, 279-285, 1991

A:Title: Distribution of plural HCV types in Japan

A:Reference number: F00554; MUID:92068204; PMID:11720309

A:Accession: F00559

A:Molecule type: mRNA

A:Residues: 2678-2729 <KAT>

A:Cross-references: GB:DI0562; GB:DI0518; NID:g221523; PIDN:BA01418.1; PID:g221524

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:116-115/Product: capsid protein C #status predicted <CPC>

F:116-131/Product: envelope protein M #status predicted <EPM>

F:139-389/Product: major envelope protein E #status predicted <NEE>

F:139-733/Product: nonstructural protein NS1 #status predicted <NS1>

F:139-1010/Product: nonstructural protein NS2 #status predicted <NS2>

F:1011-1619/Product: hepatitis virus #status predicted <NS3>

F:1234-1241/Region: nucleotide-binding motif A (P-loop)

F:1316-1321/Region: nucleotide-binding motif B

F:1320-1323/Region: DEXH motif

F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4a>

F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 65.4%; Score 1158; DB 1; Length 3033;

Best Local Similarity 67.3%; Pred. No. 1.8e-89;

Matches 206; Conservative 47; Mismatches 53; Indels 0; Gaps 0;

QY 16 AGITVFYFVAQGLIRACMLVRKAAGHYVQMAFMKLAALGTGYVDHLTPIDQMAHAG 75

DB 908 ASLIRIFVFAHALLRMCNTVRHLAAGRYQWVLLALGRMTGYIYDHLTPMSDMAAG 967

QY 76 LRLIAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARRREILLGANDPEQGM 135

DB 968 LRLIAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARRREILLGANDPEQGM 1027

QY 136 RLAPITVYSGQRTGLGCIITSLTGRDKQVGEVQVSTATQSFATCVNGVCMVTFH 195

DB 1028 RLAPITVYSGQRTGLGCIITSLTGRDKQVGEVQVSTATQSFATCVNGVCMVTFH 1087

QY 196 GAGSKTLAAGPKGPIITQWYTNVDQVLGWAQPPGARSMTPTCGSSDLVYTRHADVI 255

DB 1088 GAGSKTLAAGPKGPIITQWYTNVDQVLGWAQPPGARSMTPTCGSSDLVYTRHADVI 1147

QY 256 RGRSGSLSPRPVSYLKSGSGGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMET 315

DB 1148 RGRSGSLSPRPVSYLKSGSGGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMET 1207

QY 316 TMRTSS 321

DB 1208 TMRTSS 1213

# RESULT 12

T08841

polyprotein - douroucouli hepatitis GB virus A

C:Species: douroucouli hepatitis GB virus A

C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: T08841

R:Berker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.

J. Gen. Virol. 79, 41-45, 1998

A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.

A:Reference number: Z16486; MUID:98120818; PMID:9460920

A:Accession: T08841

A:Molecule type: translated from GB/EMBL/DDBJ

A:Status: translated from GB/EMBL/DDBJ

A:Residues: 1-3005 <EMK>

A:Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AA040502.1; PID:g2828600

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: polyprotein

Query Match 22.5%; Score 398.5; DB 2; Length 3005;

Best Local Similarity 35.4%; Pred. No. 5.3e-25;

Matches 97; Conservative 45; Mismatches 125; Indels 7; Gaps 5;

QY 49 AFMKLAALGTGYVDHLTPIDQMAHAGRLAAVEPVIFSDMEVKIITWGAADTAACGDI 108

DB 887 AFVRRLERGGVTLFQCGQVSGAAALIXDLGVALLPEVSVAROCYIYRDARLACQR 946

QY 109 ISGLPVSARRREILLG--PADNFEQGMRLAPITVYSGQRTGLGCIITSLTGRDKQ 166

DB 947 ISGLPVSARRREILLG--PADNFEQGMRLAPITVYSGQRTGLGCIITSLTGRDKQ 1005

QY 167 VEGEVQVSTATQSFATCVNGVCMVTFHGAAGSKTLAAGPKGPIITQWYTNVDQVLGWAQ 226

DB 1006 HEGSIVLTCTSTREMGTCVNGVCMVTFHGAAGSKTLAAGPKGPIITQWYTNVDQVLGWAQ 1065

QY 227 PARSMTPTCGSSDLVYTRHADVIYVRRRSGSLSPRPVSYLKSGSGGPLCPGSHAV 286

DB 1066 SGASGLPECKCGTQSWCIRN--DGALCHGRSLKVLVDLPLEISDFRSGSGSPILDCBG 1123

QY 287 HAVGIFRAVCTRGVAKAVDFIPVESMETTMTS 320  
 Db 1124 HVGGMV-VSVLARGV-KVTGVRVYKPKMETLFXDS 1155

## RESULT 13

108839  
 Polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: 108839  
 R/Enter: U.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: 108839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <SRK>  
 A/Cross-references: EMBL:AF023424; NID:92828597; PIDN:AA040501.1; PID:92828598  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 19.4%; Score 343; DB 2; Length 2970;  
 Best Local Similarity 30.5%; Pred. No. 2.7e-20;

Matches 65; Conservative 44; Mismatches 94; Indels 56; Gaps 9;

QY 71 WAHAG-----LRDLAVAVEPVSFSDMEVKITMGADTAACGDIISGLPVSARRGRE 121  
 Db 691 YAHAGVTRKTAEDLRQMGPRLEPVAVHPEDCAMVRAATLSCGSGVHGKPVVARRGDE 950  
 QY 122 ILGPDNFEQGGWRL-----LAPITAYSQQTRGLGCIITSLTGRDKNQVEGEVQVS 175  
 Db 951 VLIGVLNGV-----WELPPGFVPTAPVAVH-HHGKFGGVKTSMTGMDTEHGVNVVLG 1005  
 QY 176 TATGSPLATGVNCGVCFHFGAGSKTLGKRPITQMTYVDDLVGMQAPGARSMTPC 225  
 Db 1006 TSTTRSGTGVNCGVMTYTHGSSNARLTAAQMGVNSRWASDVAVYPLPVAKKCLEPC 1065  
 QY 236 TCGSSDLYLVRHADVIPVRRRGRSGSLSS-----PRVSYLKSSSGGFLCP 284  
 Db 1066 KCGQGVWV-----RND--GALCHGTIGRTVELDLAELCDFRSGSGSFLCD 1112  
 QY 285 SGHAGVIFRAVCTRGVAKAVDFIPVESMETTMTSSAW 323  
 Db 1113 EGHAVGML-LSVLRG-----SRVTGIRYTKFW 1139

## RESULT 14

887392  
 conserved hypothetical protein CC1155 [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: G87392  
 R/Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.D.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjol  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: G87392  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-353 <SNO>  
 A/Cross-references: GB:AE005673; NID:G13422473; PIDN:AAK23139.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC1155

Query Match 6.1%; Score 108; DB 2; Length 353;  
 Best Local Similarity 21.9%; Pred. No. 0.16;  
 Matches 77; Conservative 36; Mismatches 122; Indels 116; Gaps 14;

QY 22 PYFVRAQGLIRACMLVEXAA-----GHHV-----OMAFKLAALTGTYYVHDL 65  
 Db 65 PIVVLAGLPAFSQLRBSAIVAMRASGISGRITGMVPAVAVVLDALCGVLAAPRA 124  
 QY 66 TP-LQDMAHAGLDLAVAEPIVPSDMEVKITMGADTAACGDIISGLPVSARRGREILL 124  
 Db 125 DPLTADM-WRNTTPVAREKEPVPRTPRAGADLVIGANNAADRKITLTVITFRDSGLIV 163  
 QY 125 ----GPADNFEQGGWRLAPITAYSQQTRGLGCIITSLTGRDKNQVEGEVQVSATQS 180  
 Db 184 EKVFAAPARYDDKAWLTLEQPKT-----TRFADLDAQASTPA-- 219  
 QY 181 FLATGVNCGVMTVFGAGSKTLGKRPITQMTYVDDLVGMQAPGARSMTPTCGSS 240  
 Db 220 ----ATSWP-----TALRPQDVGLFGDDSVFTAS----- 246  
 QY 241 DLYLVRHADVIPVRRRGRSGSLSPRPVSY-----LKGSSGGP-----LTCFSGHAYG 290  
 Db 247 -----ARRALENGG---SDRPESFYATHLQAFASPVSILWLLLSAPVALA 290  
 QY 291 IFR-----AAVCTRGVAKAVDFIPVESMETTMTSSA-----WRHPQRCG 330  
 Db 291 NFRSGGAVLLTGLGACGMFLVANGMLTALDESGALTPFLAVMAAPALFG 341

## RESULT 15

B46642  
 DNA-directed DNA polymerase (BC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chair  
 C/Species: Mus musculus (house mouse)  
 C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
 C/Accession: B46642  
 R/Miyazawa, H.; Izumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanacka, F.  
 J. Biol. Chem. 268, 8111-8122, 1993  
 A/Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase  $\alpha$   
 A/Reference number: A46642; MUID:93216788; PMID:8463324  
 A/Accession: B46642  
 A/Status: preliminary  
 A/Molecule type: mRNA; protein  
 A/Residues: 1-600 <MTY>  
 A/Cross-references: GB:D13546; NID:9303658; PIDN:BA02746.1; PID:9303659  
 A/Experimental source: FMA3 cells  
 A/Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBI:129149)  
 C/Keywords: nucleotidyltransferase

Query Match 5.7%; Score 101; DB 2; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 1.3;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 30 LIRACMLVRKAAGHYVGM-AFMKLAALT-----GTYVVDHL-----TFLQWMA 72  
 Db 27 LAELCVLRQTEDEGWSVSLIAFCTSAKTCITVDILNFEYEVNKKLSKAMWSKXSG 86  
 QY 73 HAGLDLAVAEPIVPSDMEVKITMGADTAACGDI--ISGLP-----VSARRGREI 122  
 Db 87 HAGTDDI-VSIQELLEAREBEETLLSTYTSKGLKRVSSSTPEPLTKRVAARSFQD- 144  
 QY 123 LIGPDNFEQGGWRLAPITAYSQQTRGLGCIITSLTGRDKNQVEGEVQVSATQSFL 182  
 Db 145 LLSFSS-----FSPSATPSQK-----YFSRTNR-----GEVVTTFGSAQ--- 178  
 QY 183 ATCVNCGVMTVFGAGSKTL--AGPKGPIQMTYVDDLVG 222  
 Db 179 -----GLMSGRGSGSVSLKVVGDPEPLTGSYRAMFQQLNG 215

Search completed: May 6, 2004, 09:37:14  
 Job time : 12.8777 secs





Fri May 7 13:37:08 2004

us-10-650-585-4.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.97286 Seconds

(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-4

Perfect score: 1771

Sequence: 1 MKKKKLEHHHHHTSAGITK.....TMTSSAMWHPQGGKKKK 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1533	86.6	3010	1	POLG_HCVUT
2	1528	86.3	3010	1	POLG_HCVUT
3	1515	85.5	3010	1	POLG_HCVUT
4	1479	83.5	3010	1	POLG_HCVBK
5	1399	79.0	3011	1	POLG_HCVL
6	1386	78.3	3011	1	POLG_HCVL
7	1173	66.2	3033	1	POLG_HCVU6
8	1158	65.4	3033	1	POLG_HCVU6
9	101	5.7	600	1	DPO2_MOUSE
10	95.5	5.4	706	1	TRFE_HORSE
11	93	5.3	660	1	VST2_HEYBU
12	93	5.3	660	1	VST2_HEYPA
13	91.5	5.2	1705	1	PTPV_MOUSE
14	90	5.1	659	1	VST2_HEYME
15	88	5.0	3414	1	POLG_TBEVH
16	87	4.9	3412	1	POLG_TBEVS
17	87	4.9	3414	1	POLG_TBEVH
18	86.5	4.9	470	1	NRAM_IAMW
19	86	4.9	434	1	TOLB_CHLTH
20	85	4.8	470	1	NRAM_IATRA
21	85	4.8	730	1	HELS_METWA
22	84.5	4.8	347	1	MDHM_EUCGU
23	84	4.7	309	1	UCP2_PAT
24	84	4.7	339	1	GPDA_COREF
25	84	4.7	470	1	NRAM_IARUE
26	83.5	4.7	538	1	DAC_ACTSP
27	83.5	4.7	854	1	FMP2_SCHPO
28	83.5	4.7	1399	1	RPOC_PSEAE
29	83	4.7	309	1	UCP2_HUWAN
30	83	4.7	485	1	VST2_HEYVA
31	83	4.7	485	1	VST2_HEYVA
32	82.5	4.7	453	1	NRAM_IAMWL
33	82	4.6	309	1	UCP2_MOUSE

#### ALIGNMENTS

RESULT 1	ID	POLG_HCVUT	STANDARD;	PRT; 3010 AA.
AC	Q00269;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [contains: Capsid protein C (Core protein) (P22);			
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2			
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)			
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)			
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein			
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein			
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)1.			
OS	Hepatitis C virus (isolate HC-UT) (HCV).			
OC	Viruses; serina positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepatitisin.			
OX	NCBI_TaxId=31642;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92295714; PubMed=1318627;			
RA	Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.Y.,			
RA	Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;			
RT	"Molecular cloning of hepatitis C virus genome from a single Japanese			
RT	carrier: sequence variation within the same individual and among			
RT	infected individuals."			
RL	Virus Res. 23:39-53(1992).			
CC	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are			
CC	hydrophobic, suggesting a possible membrane-related function. NS3			
CC	and NS5 may play a role in the viral RNA replication.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral			
CC	precursor polyprotein, commonly with Asp or Glu in the P6			
CC	position. Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +			
CC	{RNA} (N).			
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a			
CC	lipoprotein envelope. The envelope consists of two proteins:			
CC	protein M and glycoprotein E. The nucleocapsid is a complex of			
CC	protein C and RNA.			
CC	-1- SIMILARITY: THE PROTEIN BELONGS TO PEPTIDASE FAMILY S29.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sdb.ch">license@sdb.ch</a> ).			
CC	-----			
DR	EMBL, D11168; BAA01943.1; -			
DR	PIR, A45573; A45573.			
DR	MEPROS; S29.001; -			
DR	MEPROS; U39.001; -			
DR	InterPro: IPR009003; Cys_Ser_trypsin.			
DR	InterPro: IPR001410; DEAD.			
DR	InterPro: IPR002522; HCV_capsid.			

Query Match	Best Local Similarity	Score 1533; DB 1;	Length 3010;
Matches 289; Conservative	95.1%;	Pred. No. 9, 9e-124;	Indels 0; Gaps 0
	7;	Mismatches 8;	

16 AGITKVPYVRACQGLIRACMTVRKAGGVYMAFMFKALALGTYYVYDHLPLQPMWANG 75

Db	504	AATTAMPYFPRAAGILRACMLVAKVAGHYVQVAFKKLAAALGTYYVYDHLTFLQDWAHG	963
Qy	76	LRDLAAVEVEVIFSDMEVKKITWGADITPAACGDIISGLPYISARGRFILLGPADNFEQGM	135
Db	964	LRDLAAVEVEVIFSDMEVKKITWGADITPAACGDIISGLPYISARGRFILLGPADNFEQGM	1023
Qy	136	RLAPLTAAYAOQRRLGLGCIITSLTRDKNVQGEVQVASTAQSLATCAVGVCTVH	195
Db	1024	RLAPLTAAYAOQRRLGLGCIITSLTRDKNVQGEVQVASTAQSLATCAVGVCTVH	1083
Qy	196	GAGSKTLAGEKGPITQMTVWVDDLVGMQAPFGARSMTPTCGSSDLYLVTRADYIPVR	255
Db	1084	GAGSKTLAGEKGPITQMTVWVDDLVGMQAPFGARSMTPTCGSSDLYLVTRADYIPVR	1143
Qy	256	RGGRSRSLSPRVSYLXGSSGGPILCEGHAIVGIFRAVCRGVAKAVDFIPVESMET	315
Db	1144	RGGRSRSLSPRVSYLXGSSGGPILCEGHAIVGIFRAVCRGVAKAVDFIPVESMET	1203
Qy	316	TMRT 319	
Db	1204	TMRS 1207	
RESULT 2			
POLG_HCVUA	STANDARD;	PRT; 3010 AA.	
AC	P26652;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Genome polyprotein [Contains: Capsid protein, C (core protein) (P22);		
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2		
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)		
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)		
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein		
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein		
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].		
OS	Hepatitis C virus (isolate Japanese) (HCV).		
OS	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepadnavirus.		
OX	NCBI_TaxID=11116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=9108650; Pubmed=2175903;		
RA	Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,		
RA	Sugimura T., Shimotohno K.;		
RT	"Molecular cloning of the human hepatitis C virus genome from		
RT	Japanese patients with non-A, non-B hepatitis.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).		
RN	[2]		
RP	DISCUSSION OF SEQUENCE.		
RX	MEDLINE=91192160; Pubmed=1849486;		
RA	Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,		
RA	Ohkoshi S., Shimotohno K.;		
RT	"Molecular structure of the Japanese hepatitis C viral genome.";		
RT	FEBS Lett. 280:325-328(1991).		
CC	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are		
CC	hydrophobic, suggesting a possible membrane-related function. NS3		
CC	and NS5 may play a role in the viral RNA replication.		
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral		
CC	precursor polyprotein, commonly with Asp or Glu in the P6		
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.		
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate +		
CC	{RNA} (N)		
CC	-1- SUBUNIT: The viron of this virus is a nucleocapsid covered by a		
CC	lipoprotein envelope. The envelope consists of two proteins:		
CC	protein M and glycoprotein E. The nucleocapsid is a complex of		
CC	protein C and mRNA.		
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.fsb-sib.ch/announce/>  
CC or send an email to [license@fsb-sib.ch](mailto:license@fsb-sib.ch)).

DR EMBL; D90208; BAA1433.1; -  
DR PIR; A39253; GNMVCT.  
DR HSSP; P26663; IUXP.  
DR MEROPS; S29.001; -  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002869; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RAR.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVL.  
DR Pfam; PF01543; HCV\_capsid.1.  
DR Pfam; PF01542; HCV\_core.1.  
DR Pfam; PF01539; HCV\_env.1.  
DR Pfam; PF01560; HCV\_NS1.1.  
DR Pfam; PF01538; HCV\_NS2.1.  
DR Pfam; PF02907; HCV\_NS3.1.  
DR Pfam; PF01006; HCV\_NS4a.1.  
DR Pfam; PF01001; HCV\_NS4b.1.  
DR Pfam; PF01506; HCV\_NS5a.1.  
DR Pfam; PF00271; Helicase\_C.1.  
DR Pfam; PF00998; Viral\_RAR.1.  
DR Pfam; PF0186062; HCV\_NS1.1.  
DR SMART; SMO0487; DEXDC.1.  
KM Polypeptide; Glycoprotein; Transferrin; RNA-directed RNA polymerase;  
KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.  
KM Transmembrane; Nonstructural protein; Envelope protein; Helicase; ATP-binding;  
KM Transmembrane; Nonstructural protein; Envelope protein; Helicase; ATP-binding;  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.fsb-sib.ch/announce/>  
CC or send an email to [license@fsb-sib.ch](mailto:license@fsb-sib.ch)).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;  
Query Match 86.3%; Score 1528; DB 1; Length 3010;  
Best Local Similarity 94.1%; Pred. No. 2.7e-123;  
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;  
QY 16 AGITKVPYFVRAQGLIRACMLVRAKAGHYVMAFKALALGTGVYDHLPLQDVAHAG 75  
Db 904 AGITRVYFVRAQGLIRACMLVRAKAGHYVMAFKALALGTGVYDHLPLQDVAHAG 963  
QY 76 LRDLAIVAEVPIFSMEVKKITWGADTAACGIIISGLPIASARGREIILGPDNFGQGM 135  
Db 964 LRDLAIVAEVPIFSMEVKKITWGADTAACGIIISGLPIASARGREIILGPDNFGQGM 1023  
QY 136 RLDAPIYASQQRGLIGCIITSLGRDNQVEGEYQVSTATQSFATCVAGVCTVPH 195  
Db 1024 RLDAPIYASQQRGLIGCIITSLGRDNQVEGEYQVSTATQSFATCVAGVCTVPH 1083  
QY 196 GAGSKTIAPKPGITQMTYNTVDQILVQWAPPAASMTPTCTGSSDLYVTRHADVTPVR 255  
Db 1084 GAGSKTIAPKPGITQMTYNTVDQILVQWAPPAASMTPTCTGSSDLYVTRHADVTPVR 1143  
QY 256 RGDSRGSILSPRPVYLKSGSGGPIICSGHANGIFRAAVCTRGVAKVADPIPVESMET 315  
Db 1144 RGDSRGSILSPRPVYLKSGSGGPIICSGHANGIFRAAVCTRGVAKVADPIPVESMET 1203  
QY 316 TMRK 319  
Db 1204 TMRK 1207  
RESULT 3  
POLG\_HCVTM STANDARD; PRT; 3010 AA.  
AC POLG\_HCVTM  
ID P29846;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 42, Last annotation update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Envelope glycoprotein (contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)1.  
OS Hepatitis C virus (isolate Taiwan) (HCV)  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31645;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9220206; PubMed=1314449;  
RA Chen P.-J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
RT "The Taiwanese hepatitis C virus genome: sequence determination and  
RT mapping the 5' terminus of viral genomic and antigenomic RNA";  
RL Virology 188:102-113(1992).  
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA} (N).  
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein C and mRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.







```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 3011 AA; 327197 MW; 658C9447CE5AF9 CRC64;

Query Match 79.0%; Score 1399; DB 1; Length 3011;
Best Local Similarity 84.2%; Pred. No. 3,66-112;
Matches 256; Conservative 25; Mismatches 23; Indels 0; Gaps 0;

QY 16 AGITKVPYFVAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYDHLPLDPAHAG 75
DB 904 ASLKVPEFVAVQGLRFCALARKNIGHYVQWYIKLGALTGYVNHLPFLDMANG 963
QY 76 LRDLAFAVEPIFSDMEYKLTWGAADTAACGDIISGLPVARSREILLGPADNFEQGW 135
DB 964 LRDLAFAVEPIFVQSMETKLTWGAADTAACGDIISGLPVARSREILLGPADNFEQGW 1023
QY 136 RLAPITAYSOQTRGLGCLITSLTGRDNQVEGEVQVSTATQSPFLATCVNGVQWTFH 195
DB 1024 RLAPITAYAQQTGLGCLITSLTGRDNQVEGEVQVSTATQSPFLATCVNGVQWTFH 1083
QY 196 GAGSKTIAGPCTPTOMTYNTDPLVGMQAPGARSMPTCCGSSDLYLVTRHADVPVR 255
DB 1084 GAGSKTIAGPCTPTOMTYNTDPLVGMQAPGARSMPTCCGSSDLYLVTRHADVPVR 1143
QY 256 RRGDSRGSLSPRPVSYIKSSGGPDLCPGSHAVGIFRAAVCTRGVAKAVDFPESMET 315
DB 1144 RRGDSRGSLSPRPISYIKSSGGPDLCPGSHAVGIFRAAVCTRGVAKAVDFPESMET 1203
QY 316 TMRT 319
DB 1204 TMRS 1207

RESULT 6
POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
AT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

```

```

RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467 (1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=96154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100 (1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M67463; AAA4534.1; -
DR PIR; A36814; GNVACH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AAR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR INTERPRO; IPR009003; Cys_Ser_trypsin.
DR INTERPRO; IPR001410; DEAD.
DR INTERPRO; IPR002522; HCV capsid.
DR INTERPRO; IPR002521; HCV core.
DR INTERPRO; IPR002519; HCV env.
DR INTERPRO; IPR002511; HCV NS1.
DR INTERPRO; IPR002518; HCV NS2.
DR INTERPRO; IPR000745; HCV NS4A.
DR INTERPRO; IPR001490; HCV NS4B.
DR INTERPRO; IPR002868; HCV NS5A.
DR INTERPRO; IPR002166; HCV RdRp.
DR INTERPRO; IPR001650; Helicase_C.
DR INTERPRO; IPR004109; Peptidase_C29.
DR INTERPRO; IPR007095; RNA_pol_DS_PS.
DR INTERPRO; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.

```





CC Hepacivirus.  
 OX NCBI\_TaxID=11113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93044440; PubMed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,  
 RA Machida A., Miyakawa Y., Nayumi M.,  
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704 (1991).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).  
 FT CHAIN 101 101 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 734 1010 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1619 1619 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 1867 2017 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT CHAIN 2018 3033 POTENTIAL.  
 FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1169 1169 ATP (POTENTIAL).  
 FT NR BIND 1234 1241 DECH BOX.  
 FT SITE 1320 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3033 AA; 329165 MW; P957F5C1A273B5B5 CRC64;

Query Match 66.2%; Score 1173; DB 1; Length 3033;  
 Best Local Similarity 69.2%; Pred. No. 1.2e-92;  
 Matches 209; Conservative 45; Mismatches 48; Indels 0; Gaps 0;

QY 18 ITKVPYFVARAGLIRPCMVKRAAGHYVOMFMKLAALGTGYVDHLPLDGMHAGIR 77  
 DB 910 LTRVPEFVRHMLNRCTVWRHLAGRYOMVLLAGRTGYIYIDHLPMSGMANGIR 969  
 QY 78 DLAVAVEPIVFSDFMEKIIITWGDITACGDIISGLVPSARRREIILGPADNFEQGMEL 137  
 DB 970 DLAVAVEPIVFSDFMEKVIWGAETACGDIILHGPVSRRLREVLGPADGYTSKGMEL 1029  
 QY 138 LAPITAYSOOTRGLGLCITISLTGRDKNVBESEVQVVSATATSPLATCNQVCWYVTHGA 197  
 DB 1030 LAPITAYAOQTGGLIGTIVSMGRDKTEQABEIQVLSVTQSPFGTTISGLVMTVYHGA 1089  
 QY 198 GSKTLAGPGPITOMVTNVDOLVGWQADPGARSMPTCCGSSDYLVTVRHADVIFVRR 257  
 DB 1090 GSKTLAGSGPVTOMVSSAGDVLGWPSPGKSLKPCGAVDLYIVTRMADVIFARRR 1149  
 QY 258 GDSRGSLLSPRPVSYSKSSGGGLLCPSSGHAVGIFRAAVCTGVAKADVIFVESMETTM 317  
 DB 1150 GDSRGSLLSPRPVSLKSSGGGGLLCPSSGHAVGIFRAAVCTGVAKADVIFVESMETTM 1209  
 QY 318 RT 319  
 DB 1210 RS 1211

RESULT 8  
 POLG\_HCV8 STANDARD; PRT: 3033 AA.  
 ID POLG\_HCV8  
 AC P2661;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide (contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus).

Query Match	Best Local Similarity	Matches	Conservative	Score	DB 1	Length	3033
16	AGITVPEVFAAGLIRACMLVRKAGAHVQVMAFNKLAALTGYYVYDHLRPLQDMAHAG	67.3%	47	1158	DB 1	3033	
908	ASLMLIPFVFAHMLRVCVLKHLAAGRYIOWMLITIGRMGTGIYIDHLSPLSTMAAG	67.3%	53	1158	DB 1	3033	
76	LNDLAVANPEVPIFSDMEVKITWAGDPAAGDIIISGLPSARSARPEILLGPDNDEGGW	67.3%	53	1158	DB 1	3033	
968	LNDLAIAEVPAVFSMEKRYIWAEEVAGDILHGLPVASARGREVLGPDAGITSGW	67.3%	53	1158	DB 1	3033	
136	RLIAPITVYSOOTRGLGCIITSLTGRDKNOVGEVQVWSTATQSHLACVNGVCMVFR	67.3%	53	1158	DB 1	3033	
1028	KILAIITATYTOOTRGLGAIYVSLTGRDKNEQAGVQVIVSLVQFLGHSIGVAMTYH	67.3%	53	1158	DB 1	3033	
196	GAGSKTLAAGPKPITOMTANVDOLVGMQAPFGASRNTFCTGSSSLVLYVTRHADVIVR	67.3%	53	1158	DB 1	3033	
1088	GAGNTLTLGPKCPVQWITSABGDLVGMPPSPGKSLDPTCGADVLLVYTRNAVIVR	67.3%	53	1158	DB 1	3033	
256	RGDSRGSLSPRPVSYKSSGGGLLCPSPGHVAGFRAVACITRGAQKAVDFIPVSMET	67.3%	53	1158	DB 1	3033	
1148	KQDNRGALISFRPLSTLKSSGGPVLCSRGHAGVLFRAVACRGVAKSIDIPVPSLDV	67.3%	53	1158	DB 1	3033	
316	TWRTSS 321	67.3%	53	1158	DB 1	3033	
1208	ATRTPS 1213	67.3%	53	1158	DB 1	3033	

```

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
GN POLA2..
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
OX NCBI_Taxid=10090.
RN [1]
R1 SEQUENCE FROM N. A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
R2 EX MEDLINE=99216788; PubMed=8463324.
R3 RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
R4 Hanaoka F.;
R5 "Molecular cloning of the cDNAs for the four subunits of mouse DNA
R6 polymerase alpha-primase complex and their gene expression during
R7 cell proliferation and the cell cycle.";
R8 J. Biol. Chem. 268:8111-8122(1993).
R9
R10 -1- FUNCTION: May play an essential role at the early stage of
R11 chromosomal DNA replication by coupling the polymerase
R12 alpha/primase complex to the cellular replication machinery (By
R13 similarity).
R14 -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
R15 (subunits A, B, C and D), which is assembled throughout the cell
R16 cycle. The largest subunit (subunit A) has DNA polymerase
R17 activity, the two smallest subunits (subunits C and D) have DNA
R18 primase activity. Subunit B binds to subunit A.
R19 -1- SUBCELLULAR LOCATION: Nuclear.
R20 -1- PFM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
R21 PHASE (BY SIMILARITY).
R22 -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
R23
R24 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
R25 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
R26 CC the European Bioinformatics Institute. There are no restrictions on its
R27 CC use by non-profit institutions as long as its content is in no way
R28 CC modified and this statement is not removed. Usage by and for commercial
R29 CC entities requires a license agreement (See license@isb-sib.ch).
R31 CC
R32 CC
R33 DR EMBL; D13546; BA602746.1; -.
R34 DR PIR; B46642; B46642.
R35 DR MGD; MGI:99690; Pola2.
R36 DR InterPro: IPR007200: DNA_pol_alpha_B.
R37
R38 DR Pfam; PF04058; DNA_pol_alpha_B_1.
R39
R40 DR DNA replication, Nuclear protein, Phosphorylation.
R41
R42 KW DOMAIN 101 107 POLY-GLU.
R43 FT DOMAIN 115 157 PRO/SER/THR-RICH (HYDROPHILIC).
R44
R45 SQ SEQUENCE 600 AA; 66267 MW; 79F94BEEF33FEBC CRC64;
R46
R47
R48
R49
R50
R51
R52
R53
R54
R55
R56
R57
R58
R59
R60
R61
R62
R63
R64
R65
R66
R67
R68
R69
R70
R71
R72
R73
R74
R75
R76
R77
R78
R79
R80
R81
R82
R83
R84
R85
R86
R87
R88
R89
R90
R91
R92
R93
R94
R95
R96
R97
R98
R99

```

```

DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal
binding globulin) .
GN 1o.
OS Equus caballus (Horse) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus .
CX NCBI_TaxId=9796;
RN RP SEQUENCE FROM N.A.
RA MEDLINE=93277958; PubMed=8504171;
RA Carpenter M.A., Broad T.E.;
RT "The cDNA sequence of horse transferrin." ;
RL Biochim. Biophys. Acta 1173:230-232(1993) .
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraembryonic tissue;
RA McDowell K.O., Adams W.H., Baker C.B.;
RA Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Transferrins are iron binding transport proteins which
CC can bind two atoms of ferric iron in association with the binding
CC of an anion, usually bicarbonate. It is responsible for the
CC transport of iron from sites of absorption and heme degradation to
CC those of storage and utilization. Serum transferrin may also have
CC a further role in stimulating cell proliferation.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -1- DOMAIN: Composed of two homologous domains.
CC -1- SIMILARITY: Belongs to the transferrin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch) .
CC -----
DR EMBL; M69020; AAA30958.1; -
DR EMBL; U21127; AAA63684.1; -
DR PIR; S33761; S33761.
DR HSBP; P02787; IABE.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PRO0422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN 1; 2.
DR PROSITE; PS00206; TRANSFERRIN 2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR KX Transferrin; Iron transport; Glycoprotein; Metal-binding; Repeat;
DR Signal.
KW SIGNAL.
FT 1 19 BY SIMILARITY.
FT CHAIN 20 706 SEROTRANSFERRIN.
FT REPEAT 20 357 1.
FT REPEAT 358 706 2.
FT DISULFID 26 64 BY SIMILARITY.
FT DISULFID 36 55 BY SIMILARITY.
FT DISULFID 134 215 BY SIMILARITY.
FT DISULFID 174 190 BY SIMILARITY.
FT DISULFID 177 198 BY SIMILARITY.
FT DISULFID 187 200 BY SIMILARITY.
FT DISULFID 248 262 BY SIMILARITY.
FT DISULFID 360 623 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 376 389 BY SIMILARITY.
FT DISULFID 423 701 BY SIMILARITY.
FT DISULFID 441 664 BY SIMILARITY.
FT DISULFID 474 550 BY SIMILARITY.
FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.

```

```

FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MM; 1A0FA566C0409DBA CRC64;

Query Match 5.4%; Score 95.5; DB 1; Length 706;
Best Local Similarity 21.8%; Pred. No. 1.9;
Matches 67; Conservative 43; Mismatches 115; Indels 83; Gaps 18;

QY 48 MAFKLAALGTGY--YDHLTFLQDMAHAGRLDVAVEPVIFSDMEVKIITGA---100
DB 321 LGFRIIPAMDTWLYLGEYVT-----AIRLIRREDIPREPKD-ECKKVKCAIGH 371
QY 101 DTAACGD-IISGLPVGARRR-----ELIGPAPNEFGQWRL-----LAPITAY 144
DB 372 EKVKCDSEVNSGGINCESAOSTEDCTAKIVGEALMSLDGFTYIACKGLVPLAE 431
QY 145 SQOTRGLGCIITSLTRDKNOVEGEVOVSTATQSFATCVNGVCTVTFHAGSKTLAG 204
DB 432 NYEIRSSACVDTEEGYH-----AAVAVKSSSDPLT-----WSLKG 470
QY 205 PKGPIITQMYNNVDOLVGMQAPPGARSMTPTCGSSDLIYVTRHADVIPRRGDSRGL 264
DB 471 KK---SCHTGVDR-TAGMNI PMGL-----LYSEIKHCEPDFKFRGCAKPGYA 513
QY 265 LSPRPVYLYKSSSGCP-LLC-PSGHA-----VGIFRAVCTRGVAKAVDFIPVSMSE--T 315
DB 514 RNSLTLCLICISAGCPRECEPNHERRYGTAFGLVEKGDVA---FVKHQIYEQNT 569
QY 316 TMRSSAW 323
DB 570 DGRNPDM 577

RESULT 11
VST2_HEVBU STANDARD; PRT; 660 AA.
ID VST2_HEVBU
AC P29326;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, last sequence update)
DT 01-FEB-1994 (Rel. 28, last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=31767;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.W., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R.;
RA "Hepatitis E virus (HEV): molecular cloning and sequencing of the

```

```

RT full-length viral genome."
RL virology 185:120-131(1991).
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: W73218; AAA45736.1;
DR PIR: C40778; YHMH2.
DR InterPro: IPR004261; SP2.
DR InterPro: IPR008975; Viral_cap_coat.
DR Pfam: PF03014; SP2; 1.
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 660 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70978 MM; 5832A013CC4A61C CRC64;

Query Match 5.3%; Score 93; DB 1; Length 660;
Best Local Similarity 19.0%; Pred. No. 2.8;
Matches 72; Conservative 47; Mismatches 119; Indels 140; Gaps 16;

QY 27 AGLIRACMLVKAAGHYVGMQAPFKLAALGTGYDHLTFLQDMAHAGRLDVAVEPV 86
DB 188 AATIRYPLVFNNAVGVAISISFWPTTPTTSV-----DMNSTSTVDILVQPG 239
QY 87 IFSDMEVKIITGADTAACDIIISGLPVSAARGEILIGPAD--NFEQGWRLIAPIT- 143
DB 240 IASLVI-----PERRLHYRNQGRSVETSGVA 267
QY 144 VSQOTRGL-----GCIIITSLTG-----161
DB 268 EREATSGVLMCIHISLVNSYTNTPYTGALGLDPALEFRNLTPGNTNTRVRSYSTA 327
QY 162 --RDKNQVEGEVOVSTATQSFIA---TCVNGV-----CMTVPH-----195
DB 328 RHRLRGADGTAELTTIATRFMKDLVTSNGVGEIGRGLALFLFNADTLGLGPREL 387
QY 196 --GAG-----SKTLGAPRG-PITQMYTNVDOLVGMQAPPGARSMTPTCGSSDLIYV-- 245
DB 388 ISSAGQLFYERPVVVSANGEPVTKLYTSVENA---QODKGIALPHDIDLGSRVVIQDY 443
QY 246 --TRHADVI PVARRGDSRG-SILSPRPVSYK-----GSSGPIILCPSSHAYGIF 292
DB 444 DNQHEQDRPTSPAPSRPSVLRANDVWLSTLAETDQSTYGSSTGEVTV--SDSVTLV 501
QY 293 RAAVCTRGVAKAVDFIPV 310
DB 502 NVATGAQAVARSIDWTKV 519

RESULT 12
VST2_HEVPA STANDARD; PRT; 660 AA.
ID VST2_HEVPA
AC P33426;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, last sequence update)
DT 01-FEB-1994 (Rel. 28, last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Pakistan) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=33774;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,

```

[illegible]





```
FT STRAND 460 467
FT HELIX 468 470
FT TURN 475 476
FT STRAND 477 481
FT TURN 484 485
FT TURN 487 488
FT STRAND 492 496
FT HELIX 497 501
FT TURN 502 502
FT STRAND 507 508
FT TURN 510 511
FT STRAND 516 516
FT TURN 517 521
FT HELIX 518 521
FT STRAND 522 522
FT STRAND 528 528
FT TURN 529 530
FT TURN 531 531
FT STRAND 534 536
FT STRAND 541 547
FT HELIX 548 550
FT TURN 553 557
FT STRAND 558 559
FT TURN 560 562
FT STRAND 567 573
FT TURN 575 576
FT STRAND 577 577
FT TURN 582 583
FT STRAND 586 586
```

Query Match 5.0%; Score 88; DB 1; Length 3414;

Best Local Similarity 23.0%; Pred. No. 59;

Matches 59; Conservative 28; Mismatches 86; Indels 84; Gaps 12;

```
QY 64 HLEPLD-----WAHAGRLDVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVA 116
DB 1442 HLEPLD-----WAHAGRLDVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVA 1496
QY 117 RRGREILLGPADNFEQGWRLAPITAYSOOTRGLGCIITSLGRDKNOVGEVQVST 176
DB 1497 RRGREILLGPADNFEQGWRLAPITAYSOOTRGLGCIITSLGRDKNOVGEVQVST 1532
QY 177 ATQSFATCVNGVWTFHAG---SKTLGPKGPITQWTVNTVDOLV-----GM 223
DB 1533 GSKGVLT-----NMHTRGALSIDDAVAGP-----YADVREDVVCYGAMSLDEKN 1581
QY 224 QA-----PPGARSMTPCTCGSSDLVLTNRHADVIPIRRGDSRGSLSPPRVASYLK 275
DB 1582 KGETVOYHAPPG-RAHEVHQCQGEILDT-----GRKLGAIPIDLVK 1625
QY 276 SSGGPLLCPSGHAYGIF 292
DB 1626 TSGSPILNAOGVWVGLY 1642
```

Search completed: May 6, 2004, 09:31:48  
Job time : 9.97286 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 31.2384 Seconds  
(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKLHHHHHNSAGITK.....TWTSSAMHPQGGKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1551	87.6	3010	12 Q9J3H7	Q9J3H7 hepatitis c
2	1546	87.3	3010	12 Q68826	Q68826 hepatitis c
3	1546	87.3	3010	12 P90191	P90191 hepatitis c
4	1545	87.2	3010	12 Q9DTE6	Q9DTE6 hepatitis c
5	1543	87.1	3010	12 Q9DTE4	Q9DTE4 hepatitis c
6	1542	87.1	3010	12 Q9DTE6	Q9DTE6 hepatitis c
7	1541	87.0	3010	12 P88803	P88803 hepatitis c
8	1540	87.0	3010	12 Q9J3H5	Q9J3H5 hepatitis c
9	1538	86.8	3010	12 Q9J3F9	Q9J3F9 hepatitis c
10	1535	86.7	3010	12 Q9J3F4	Q9J3F4 hepatitis c
11	1535	86.7	3010	12 Q9J3H3	Q9J3H3 hepatitis c
12	1535	86.7	3010	12 Q9J3H2	Q9J3H2 hepatitis c
13	1534	86.6	3010	12 Q9J3H0	Q9J3H0 hepatitis c
14	1534	86.6	3010	12 Q9J3H0	Q9J3H0 hepatitis c
15	1533	86.6	3010	12 Q9J3H3	Q9J3H3 hepatitis c
16	1533	86.6	3010	12 Q9J3H3	Q9J3H3 hepatitis c

17	1532	86.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
18	1532	86.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
19	1532	86.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
20	1532	86.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
21	1532	86.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
22	1532	86.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
23	1531	86.4	3010	12 Q9J3H7	Q9J3H7 hepatitis c
24	1531	86.4	3010	12 Q9J3H7	Q9J3H7 hepatitis c
25	1531	86.4	3010	12 Q9J3H7	Q9J3H7 hepatitis c
26	1530	86.4	3010	12 Q9J3H7	Q9J3H7 hepatitis c
27	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
28	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
29	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
30	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
31	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
32	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
33	1529	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
34	1528	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
35	1528	86.3	3010	12 Q9J3H7	Q9J3H7 hepatitis c
36	1527	86.2	3010	12 Q9J3H7	Q9J3H7 hepatitis c
37	1527	86.2	3010	12 Q9J3H7	Q9J3H7 hepatitis c
38	1527	86.2	3010	12 Q9J3H7	Q9J3H7 hepatitis c
39	1526	86.2	3010	12 Q9J3H7	Q9J3H7 hepatitis c
40	1525	86.1	3010	12 Q9J3H7	Q9J3H7 hepatitis c
41	1524	86.1	3010	12 Q9J3H7	Q9J3H7 hepatitis c
42	1524	86.1	3010	12 Q9J3H7	Q9J3H7 hepatitis c
43	1524	86.1	3010	12 Q9J3H7	Q9J3H7 hepatitis c
44	1524	86.1	3010	12 Q9J3H7	Q9J3H7 hepatitis c
45	1523	86.0	3010	12 Q9J3H7	Q9J3H7 hepatitis c

# ALIGNMENTS

RESULT 1  
ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.  
AC Q9J3H7;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DS Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitis C virus.  
OC NCBI\_Taxid=1103;  
RN [1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=MDIS;  
RA Nagayama K., Kuroski M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease progression."  
RT Submitted (Nov-1999) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL; AF207756; AAF65946.1; -  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR PIR; P00804; P00804.  
DR PIR; P80329; P80329.  
DR HSP; P26663; IUXP.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro; IPR000345; Cys\_Ser\_lypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR004109; peptidase C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR ProDom; PD186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Query Match 87.6%; Score 1551; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 4.6e-126;  
 Matches 290; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

DR 16 AGITKVFYFVAQGLIRACMLVRKAGHYVVAFMLAALTGTYYVDHILPIQDVAHAG 75  
 DR 904 AGITRMVYFVAQGLIRACMLVRKAGHYVVAFMLAALTGTYYVDHILPIQDVAHAG 963  
 DR 76 LRDLAVAVPPIFSDMEVKITITWADPAAGCDITSLPVARRGRETLGPANFEGQGM 135  
 DR 964 LRDLAVAVPPIFSDMEVKITITWADPAAGCDITSLPVARRGRETLGPANFEGQGM 1023  
 DR 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVGVVSTATQSFATCINGVCTVFFH 195  
 DR 1024 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVGVVSTATQSFATCINGVCTVFFH 1083  
 DR 196 GAGSKITLAGPKGPIITQMTYNTVDLVMQAPPGARSMTPCTCGSSDLYLTRHADVIPIVR 255  
 DR 1084 GAGSKITLAGPKGPIITQMTYNTVDLVMQAPPGARSMTPCTCGSSDLYLTRHADVIPIVR 1143  
 DR 256 RRGSRGSLSPRPVSYLKGSSGPIILCPSGHAGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DR 1144 RRGSRGSLSPRPVSYLKGSSGPIILCPSGHAGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 DR 316 TMRT 319  
 DR 1204 TMRS 1207

DR 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=U3;  
 RA Cho M.U.;  
 RT "Molecular cloning of Hepatitis C virus genome from a single Japanese  
 patient.";  
 RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; D14484; BA03575.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; P0246; P0246.  
 DR PIR; P0804; P0804.  
 DR PIR; P80329; P80329.  
 DR HSRP; P26663; LUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; peptidase C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR ProDom; PD186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydrolyase; Nonstructural protein; Polyprotein;  
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327150 MW; 7270F47984554FAD CRC64;

Query Match 87.3%; Score 1546; DB 12; Length 3010;  
 Best Local Similarity 96.1%; Pred. No. 1.3e-125;

Matches 292; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITVPPFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAG 75  
 DB 904 AGITAVPFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAG 963  
 QY 76 LRDIAVAEVPVFSMEKIIITWGDITACGDIISGLVPSARRGREIILGPADNFEQGW 135  
 DB 964 LRDIAVAEVPVFSMEKIIITWGDITACGDIISGLVPSARRGREIILGPADNFEQGW 1023  
 QY 136 RLAPITAVSQOTRGLICITITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFH 195  
 DB 1024 RLAPITAVSQOTRGLICITITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFH 1083  
 QY 196 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVR 255  
 DB 1084 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVR 1143  
 QY 256 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMR 319  
 DB 1204 TMR 1207

RESULT 3  
 ID P90191 PRELIMINARY; PRT; 3010 AA.  
 AC P90191;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RC STRAIN=HCV-1b;  
 RA Enomoto N.;  
 RA "Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b";  
 RT J. Clin. Invest. 96:224-230 (1995).  
 RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL; D50482; EMBL05073.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; P00254; P00254.  
 DR PIR; P00804; P00804.  
 DR PIR; P0329; P0329.  
 DR PDB; 1DXV; 12-JAN-01.  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR Interpro; IPR009003; Cys Ser\_tryptsin.  
 DR Interpro; IPR001410; DEAD.  
 DR Interpro; IPR002522; HCV\_capsid.  
 DR Interpro; IPR002521; HCV\_core.  
 DR Interpro; IPR002519; HCV\_env.  
 DR Interpro; IPR002531; HCV\_NS1.  
 DR Interpro; IPR002518; HCV\_NS2.  
 DR Interpro; IPR000745; HCV\_NS4a.  
 DR Interpro; IPR001490; HCV\_NS4b.  
 DR Interpro; IPR002868; HCV\_NS5a.  
 DR Interpro; IPR002166; HCV\_NS5b.  
 DR Interpro; IPR001650; Helicase\_C.  
 DR Interpro; IPR004109; Peptidase\_C29.  
 DR Interpro; IPR007095; RNA\_pol\_D5\_PS.  
 DR Interpro; IPR007094; RNA\_pol\_P5vir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXdc; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein.  
 KW Polypeptide; RNA-directed RNA polymerase; transferase; transmembrane.  
 FT CHAIN 1  
 FT CHAIN 191  
 FT CHAIN 192  
 FT CHAIN 384  
 FT CHAIN 809  
 FT CHAIN 810  
 FT CHAIN 1026  
 FT CHAIN 1027  
 FT CHAIN 1657  
 FT CHAIN 1658  
 FT CHAIN 1711  
 FT CHAIN 1712  
 FT CHAIN 1972  
 FT CHAIN 1973  
 FT CHAIN 2419  
 FT CHAIN 2420  
 FT CHAIN 3010  
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 87.3%; Score 1546; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 1,3e-125;  
 Matches 291; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 16 AGITVPPFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAG 75  
 DB 904 AGITAVPFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAG 963  
 QY 76 LRDIAVAEVPVFSMEKIIITWGDITACGDIISGLVPSARRGREIILGPADNFEQGW 135  
 DB 964 LRDIAVAEVPVFSMEKIIITWGDITACGDIISGLVPSARRGREIILGPADNFEQGW 1023  
 QY 136 RLAPITAVSQOTRGLICITITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFH 195  
 DB 1024 RLAPITAVSQOTRGLICITITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFH 1083  
 QY 196 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVR 255  
 DB 1084 GAGSKTLGPKGPIITOMTNTVDOLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVR 1143  
 QY 256 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMR 319  
 DB 1204 TMR 1207

RESULT 4  
ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
AC Q9DTE6; 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polypeptide.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HCV142;  
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
RA Mishiro S.;  
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
with hepatocellular carcinoma: the 'progression score' revisited.";  
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL; AB049091; BAB18804.1; -.  
CC PIR; A61196; A61196.  
CC PIR; PS0329; PS0329.  
DR HSP; P26663; LUXP.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006118; F:electron transport; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral transformation; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR00345; Cys\_Ser\_Trypsin.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NSI.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR000745; HCV\_NS4.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_NS5b.  
DR InterPro; IPR001850; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NSI; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRP; 1.  
DR ProDom; PD186062; HCV\_NSI; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC\_C; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KM Hydroxylase; Nonstructural protein; Polypeptide; Transmembrane.  
KM RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;  
Query Match 87.2%; Score 1545; DB 12; Length 3010;  
Best Local Similarity 95.4%; Pred. No. 1.6e-125;  
Matches 290; Conservative 9; Mismatches 5; Indels 0; Gaps 0;  
QY 16 AGITKVPFVPAAGLIRACMLVRAAGHYVQVAFMKALATGYYVDHLPDQNAHAG 75  
DB 904 AGITRVPFVPAAGLIRACMLVRAAGHYVQVAFMKALATGYYVDHLPDQNAHAG 963  
QY 76 LRDIAVAEPVIFSDMEVKIITWGDADPAAGDIIISGLPVASRRREIILGPADNFEGQGV 135  
DB 964 LRDIAVAEPVIFSDMEVKIITWGDADPAAGDIIISGLPVASRRREIILGPADNFEGQGV 1023  
QY 136 RLAPITVYSCQTRGLACITSLTGRDQVQVEGVQVSTATQSFATCNGVQVVFH 195  
DB 1024 RLAPITVYSCQTRGLACITSLTGRDQVQVEGVQVSTATQSFATCNGVQVVFH 1083  
QY 196 GAGSKTLAGPKGPTTOMTYNDQDLVGMQAPPGARSMTPCTGSSDLYLTVRHADVIPVR 255  
DB 1084 GAGSKTLAGPKGPTTOMTYNDQDLVGMQAPPGARSMTPCTGSSDLYLTVRHADVIPVR 1143  
QY 256 RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIRPAVCTRGVAKANDFTVESMET 315  
DB 1144 RRGDSRGSLSPRPVSYLKSSGGPILCPGSHAVGIRPAVCTRGVAKANDFTVESMET 1203  
QY 316 TWRT 319  
DB 1204 TWRS 1207  
RESULT 5  
ID Q9DTE4 PRELIMINARY; PRT; 3010 AA.  
AC Q9DTE4; 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polypeptide.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HCV1150;  
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
RA Mishiro S.;  
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
with hepatocellular carcinoma: the 'progression score' revisited.";  
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL; AB049093; BAB18806.1; -.  
CC PIR; A61196; A61196.  
CC PIR; P00246; P00246.  
CC PIR; P00804; P00804.  
CC PIR; PS0329; PS0329.  
DR HSP; P26663; LUXP.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.

DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0009003; Cys\_Ser\_tyrp\_synth.  
 DR InterPro: IPR000345; CysC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF02907; HCV\_NS2; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00467; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SO SEQUENCE 3010 AA; 327324 MW; 3DE6CF249BD151C CRC64;

Query Match 87.1%; Score 1543; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 2.3e-125;  
 Matches 290; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVFYFRAQGLIRACMLVRKAGHYVQMAFMALALTGYYVDHLTPIDMAHAG 75  
 Db 904 AGITKVFYFRAQGLIRACMLVRKAGHYVQMAFMALALTGYYVDHLTPIDMAHAS 963  
 QY 76 LRDLAVAVEPVTSDEMEKTIITWAGDTAACDITISGLPVASRRGREITLLPADNFEQGN 135  
 Db 964 LRDLAVAVEPVTSDEMEKTIITWAGDTAACDITISGLPVASRRGREITLLPADNFEQGN 1023  
 QY 136 RLAPITAYSQQRGLIGCITISLTGRDKQVEGEVQVVSATQSFATCVMNCVTFH 195  
 Db 1024 RLAPITAYSQQRGLIGCITISLTGRDKQVEGEVQVVSATQSFATCVMNCVTFH 1083  
 QY 196 GAGSKTAGKPGPTTYNTVNDLVGMQAPPGASMTPTCCSSDLVLTTRADVIPIVR 255  
 Db 1084 GAGSKTAGKPGPTTYNTVNDLVGMQAPPGASMTPTCCSSDLVLTTRADVIPIVR 1143  
 QY 256 RRGDSRGSLSPPRVSLKXSSGGPILCPGSHVIGIFRAVVCRRGAKXANDFIPIVSMET 315  
 Db 1144 RRGDSRGSLSPPRVSLKXSSGGPILCPGSHVIGIFRAVVCRRGAKXANDFIPIVSMET 1203  
 QY 316 TMRT 319  
 |||

Db 1204 TMRS 1207  
 RESULT 6  
 ID Q9DTH6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTH6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV221;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 Mishiro S.;  
 RT Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC EMBL: AB049101; BAB1814.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26663; IUXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005824; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0016787; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_tyrp\_synth.  
 DR InterPro: IPR000345; CysC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.

DR	Pfam: PF01001; HCV NS4B; 1.
DR	Pfam: PF01566; HCV NS5A; 1.
DR	Pfam: PF00271; helicase C; 1.
DR	Pfam: PF00998; viral RdRp; 1.
DR	ProDom: PD18662; HCV_NS1; 1.
DR	SMART; SMO0487; DEXDC; 1.
DR	PROSITE; PS00190; CYTOCHROME_C; 1.
KW	ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KM	Hydrophobic; Nonstructural protein; Polyprotein;
SV	RNA-directed RNA polymerase; Transferase; Transmembrane-
SQ	SEQUENCE 3010 AA; 327108 MW; DE182D810E78EE4 CR664;

  

Query Match	87.1%;	Score 1542;	DB 12;	Length 3010;
Best Local Similarity	95.7%;	Pred. No. 2.8e-125;		
Matches 291;	Conservative	7;	Mismatches 6;	Indels 0;
Gaps	0;			

  

Qy	16 AGITKVFYFRAAGILIRACMLVRKAAGHYVGMAFMKLAALTGYVDHLTPLODWASAG 75
Dd	904 AVLIKVEYFYRAGGLIRACMLVRKYAVGHHYVMAMTKALALTGYVDHLLPLQDAHAG 963
Qy	76 LRLDAVAEVEVISDEMEVKIITWGADPTAACGDIIISGLPYSAARRREILLGPANFEQGQM 135
Dd	964 LRDLAAVEAEVVSDDMETKITTWGADPTAACGDIISGLPVSARRREILLGPASFEGQM 1022
Qy	136 RLAPLITAYSQOTRGILGCIITSLTGRDNQVEGEHVGVSTATOSFLATCWNGVCWTVEH 195
Dd	1024 RLAPLITAYSQOTRGILGCTIVSLTGRDNQVEGEHVGVSTATOSFLATCWNGVCWTVEH 1083
Qy	156 GAGSKTLTAGRGKITOMVTNNVODLVGMCAFPGARSMTPCTCCSSDLYLVTRHADVIPR 255
Dd	1084 GAGSKTLTAGRGKITOMVTNNVODLVGMPPARGARSLTPCTCCSDLYLVTRHADVIPR 1143
Qy	256 RRGDSRSGLSPRPVSYSYLKGSSGGPLLCPSGHVGIFRAAVCTRGAKADVFIPEVSEMT 315
Dd	1144 RRGTRRSGLSPRPVSYSYLKGSSGGPLLCPSGHIGVFRAAVCTRGAKVADVFIPEVSEMT 1203
Qy	316 TMRT 319
Dd	1204 TMRS 1207

  

RESULT 7
PR8803 PRELIMINARY; PRT: 3010 AA.
ID PR8803;
AC PR8803;
DT 01-MAY-1997 (TEMBLrel_03, Created)
DT 01-MAY-1997 (TEMBLrel_03, Last sequence update)
DT 01-OCT-2003 (TEMBLrel_25, last annotation update)
Dr Genome polypotein.
Os Hepatitis C virus.
OC Hepacivirus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxId=11103;
Rv [1]
Rp SEQUENCE FROM N.A.
Rc STRAIN=HCV-1b;
Rn Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
[2]
Rp SEQUENCE FROM N.A.
Rc STRAIN=HCV-1b;
Rx MEDLINE=95340824; PubMed=7542279;
Ra Enomoto N., Sakuma I., Asahina Y., Kuroaki M., Murakami T., Yamamoto C., Izumi N., Matsumo F., Sato C.;
"Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b.";
J. Clin. Invest. 96:224-230(1995).
-1- SUBUNIT. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS. PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).
EMBL; D50484; BAAC0075.1; -.
PIR; A61196; A61196.

```

Db      1084 GAGSKTLAGPPIQMTYNNVDOLVGMQAPPGARSITPTCGSSDLVLTTHADVIPVR 1143
QY      256 RRGDSRGLSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAVCTRGVAKAVDIPVESMET 315
Db      1144 RRGDSRGLSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAVCTRGVAKAVDIPVESMET 1203
QY      316 TMRT 319
Db      1204 TMRS 1207

RESULT 8
QJ3H5 PRELIMINARY; PRT; 3010 AA.
ID QJ3H5
AC 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB Genome polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD17;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MNA (BY SIMILARITY).
CC EMBL; AF207758; AAF65948.1; -.
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00254; P00254.
DR PIR; P00329; P00329.
DR HSBP; P27958; HEB1.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F: electron transporter activity; IEA.
DR GO; GO:0016787; F: hydrolase activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR GO; GO:0016740; F: transferase activity; IEA.
DR GO; GO:0006118; F: electron transport; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR009003; Cys Ser tyrosin.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR005522; HCV_capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV_NS4.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR0041650; Helicase_C.
DR InterPro: IPR004109; peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.

```

```

DR InterPro: IPR007095; RNA_pol_PsVlr.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01558; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326801 MW; 9FE3D1B93B7AA4B CRC64;

Query Match 87.0%; Score 1540; DB 12; Length 3010;
Best Local Similarity 95.4%; Pred. No. 4.2e-125;
Matches 290; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYVRAQGLIRACMLYRKAGHYVQMAFKALATGTYVYDHLTPLQMAHAG 75
Db 904 AGITRVPYVRAQGLIRACMLYRKAGHYVQMAFKALATGTYVYDHLTPLQMAHAG 963
QY 76 LRDLAAVPEVTFPSDEVEVITWGADTAACGDIISGLPSARGRRIILGPANFEGGQM 135
Db 964 LRDLAAVPEVTFPSDEVEVITWGADTAACGDIISGLPSARGRRIILGPANFEGGQM 1023
QY 136 RLIAPIYVSGQTRGLGCIITSLTGRDNQVEGEVQVSTATQSLFATCVAGCVTVYH 195
Db 1024 RLIAPIYVSGQTRGLGCIITSLTGRDNQVEGEVQVSTATQSLFATCVAGCVTVYH 1083
QY 196 GAGSKTLAGPPIQMTYNNVDOLVGMQAPPGARSMTPTCGSSDLVLTTHADVIPVR 255
Db 1084 GAGSKTLAGPPIQMTYNNVDOLVGMQAPPGARSMTPTCGSSDLVLTTHADVIPVR 1143
QY 256 RRGDSRGLSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAVCTRGVAKAVDIPVESMET 315
Db 1144 RRGDSRGLSLSPRPVSYLKGSSGGPILCPGSHAVGIFRAVCTRGVAKAVDIPVESMET 1203
QY 316 TMRT 319
Db 1204 TMRS 1207

RESULT 9
QJ3H5 PRELIMINARY; PRT; 3010 AA.
ID QJ3H5
AC 0807P3
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DB Polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M1E;
RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Mochi T.,
RA Nio Y., Hosaka Y., Miyahara Y., Shimotohno K.;
RT "Subgenomic replicon derived from a cell line infected with the
RT hepatitis C virus."
RL Biochem. Biophys. Res. Commun. 293:993-999 (2002).
DR EMBL; AB080299; BAC54896.1; -.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.

```

DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F:electron transporter activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0003668; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR GO: GO:0006118; F:electron transport; IEA.  
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F:transcription; IEA.  
 DR GO: GO:0019079; P:viral genome replication; IEA.  
 DR GO: GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4B.  
 DR InterPro: IPR002868; HCV\_NS5A.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4; 1.  
 DR Pfam: PF01001; HCV\_NS4B; 1.  
 DR Pfam: PF00271; HCV\_NS5A; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00480; HELIC\_C; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR PolyPhen.  
 SQ SEQUENCE 3010 AA; 327097 MW; E56418C7A723B686 CRC64;

Query Match 87.0%; Score 1540; DB 12; Length 3010;  
 Best local similarity 95.7%; Pred. No. 4,2e-125;  
 Matches 291; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVAAGLIRACMLVRRAGGHVYQMAFMKLAALTGYVVDHTPLQDWAHAG 75  
 DB 904 AGITRVPEFVAAGLIRACMLVRRAGGHVYQMAFMKLAALTGYVVDHTPLQDWAHAG 963  
 QY 76 LRDIAVNEPPIESDMEVKITTWGADTAAGDIIISGLPVSARRGRETLLGPADNFEQGM 135  
 DB 964 LRDIAVNEPPIESDMEVKITTWGADTAAGDIIISGLPVSARRGRETLLGPADNFEQGM 1023  
 QY 136 RLAPITAYSQQTGLGCIITSLTGRDKQVEGEVQVSTATQSFATLVNGVCMVTFH 195  
 DB 1024 RLAPITAYSQQTGLGCIITSLTGRDKQVEGEVQVSTATQSFATLVNGVCMVTFH 1083  
 QY 196 GAGSKTLAGEPPIFTQWNTNDOLVGMQAPRGASMTPTCCGSSDLVLTTRADVTPVR 255  
 DB 1084 GAGSKTLAGEPPIFTQWNTNDOLVGMQAPRGASMTPTCCGSSDLVLTTRADVTPVR 1143  
 QY 256 RRGDSRGSLPPRPSVYLKSGSGPPLCPSGHAGVIFRAAVCTRGVAKADFIIVESMET 315  
 DB 1144 RRGDSRGSLPPRPSVYLKSGSGPPLCPSGHAGVIFRAAVCTRGVAKADFIIVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 10  
 ID Q9J3F9 PRELIMINARY; PRT: 3010 AA.  
 AC Q9J3F9  
 DT 01-OCT-2000 (TRENDELrel. 15, Created)  
 DT 01-OCT-2000 (TRENDELrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TRENDELrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD3;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease progression."  
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AF207774; AF659641; -.  
 DR PIR: A61196; A61196.  
 DR PIR: PQ0246; PQ0246.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P27958; 1HE1.  
 DR MEROPS: S29\_001; -.  
 DR MEROPS: U9\_001; -.  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0019028; C:viral capsid; IEA.  
 DR GO: GO:0019031; C:viral envelope; IEA.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F:electron transporter activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0003668; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR GO: GO:0006118; F:electron transport; IEA.  
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F:transcription; IEA.  
 DR GO: GO:0019079; P:viral genome replication; IEA.  
 DR GO: GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR00345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4B.  
 DR InterPro: IPR002568; HCV\_NS5A.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001490; HCV\_NS4B.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4; 1.  
 DR Pfam: PF01001; HCV\_NS4B; 1.  
 DR Pfam: PF01506; HCV\_NS5A; 1.



DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; viral\_RBP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SQ SEQUENCE 3010 AA, 327102 MW, 716209B93E60C7 CRC64;

Query Match 86.8%; Score 1538; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 6.4e-125;  
 Matches 288; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 75  
 DB 904 AGITRMPYFVRAQGLIRACMLVRKAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 963  
 QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 135  
 DB 964 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 1023  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCWTFPH 195  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCWTFPH 1083  
 QY 196 GAGSKTLAPKPGPTTQMTYTNVDLGVQAPPGARSMTPTCGSSDLVYTRHADVPVR 255  
 DB 1084 GAGSKTLAPKPGPTTQMTYTNVDLGVQAPPGARSMTPTCGSSDLVYTRHADVPVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMET 315  
 DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 11  
 ID 070815 PRELIMINARY; PRT; 361 AA.  
 AC 070815;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polypeptide (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98321154; PubMed=9656998;  
 RA Yamada K., Mori A., Seki M., Kimura J., Yuasa S., Matsuura Y.,  
 RA Miyamura T.,  
 RT "Critical point mutations for hepatitis C virus NS3 proteinase.",  
 RT Virology 246:104-112(1998).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,  
 RA Miyamura T.,  
 RT "Enzymatic characterization of purified NS3 serine proteinase of  
 RT hepatitis C virus expressed in Escherichia coli.",  
 RL FEBS Lett. 378:37-42(1998).  
 DR EMBL; AB013620; BAA28498.1; -  
 DR HSSP; P27958; 1HEI.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_Tyrosin.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.  
 FT NON TER 1  
 FT NON TER 361  
 SQ SEQUENCE 361 AA, 38336 MW, 870C310C76F4BC3 CRC64;  
 Query Match 86.7%; Score 1535; DB 12; Length 361;  
 Best Local Similarity 94.7%; Pred. No. 7.1e-126;  
 Matches 288; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 75  
 DB 5 AGITRMPYFVRAQGLIRACMLVRKAGHYVQMAFMKALALGTYYVDHLTPLODMAHAG 64  
 QY 76 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 135  
 DB 65 LRDLAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVSAARGREIILGPADNFEQGM 124  
 QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCWTFPH 195  
 DB 125 RLAPITAYSQOTRGLGCIITSLTGRDKQVGEVQVSTATQSFATCNGVCWTFPH 184  
 QY 196 GAGSKTLAPKPGPTTQMTYTNVDLGVQAPPGARSMTPTCGSSDLVYTRHADVPVR 255  
 DB 185 GAGSKTLAPKPGPTTQMTYTNVDLGVQAPPGARSMTPTCGSSDLVYTRHADVPVR 244  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMET 315  
 DB 245 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMET 304  
 QY 316 TMRT 319  
 DB 305 TMRS 308

RESULT 12  
 ID 0903F4 PRELIMINARY; PRT; 3008 AA.  
 AC 0903F4;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 DE MD34.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD34;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.,  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 RT progression.",  
 RT Submitted (Nov-1999) to the EMBL/Genbank/DBSI databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL; AF208024; AAF61205.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00245; P00245.  
 DR PIR; P00245; P00245.  
 DR HSSP; P26663; 1UXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005488; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

ID	AC	09J3H3;	PRELIMINARY;	PRT; 3010 AA.	AD
AC	09J3H3;				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE		Genome Polyprotein.			
OS		Hepatitis C virus.			
OC		Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC		Hepadnavirus.			
OX		NCBI_Taxid=1103;			
AP		SEQUENCE FROM N.A.			
RC		STRAIN=MD15;			
RA		Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.			
RT		"Characteristics of hepatitis C viral genome associated with disease progression.";			
RL		Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.			
CC		-1- SUBMITT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A			
CC		LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS OF			
CC		PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF			
DR	EMBL;	AF207760; AAF65950.1; -			
DR	PIR;	A61196; A61196.			
DR	PIR;	PS0329; PS0329.			
DR	HSSP;	P26663; 1xyp			
DR	GO;	GO:0016021; C:integral to membrane; IEA.			
DR	GO;	GO:0019028; C:viral capsid; IEA.			
DR	GO;	GO:0019031; C:viral envelope; IEA.			
DR	GO;	GO:0005524; F:ATP binding; IEA.			
DR	GO;	GO:0008025; F:ATP dependent helicase activity; IEA.			
DR	GO;	GO:0005489; F:electron transporter activity; IEA.			
DR	GO;	GO:0003723; F:RNA binding; IEA.			
DR	GO;	GO:0003966; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO;	GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO;	GO:0005199; F:structural molecule activity; IEA.			
DR	GO;	GO:0016740; F:transferase activity; IEA.			
DR	GO;	GO:0006118; F:electron transport; IEA.			
DR	GO;	GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO;	GO:0006350; P:transcription; IEA.			
DR	GO;	GO:0019079; P:viral genome replication; IEA.			
DR	GO;	GO:0019087; P:viral transformation; IEA.			
DR	InterPro;	IPR009003; Cys Ser tyrosin.			
DR	InterPro;	IPR000345; CytC_heme_BS.			
DR	InterPro;	IPR001410; DEAD.			
DR	InterPro;	IPR002522; HCV_capsid.			
DR	InterPro;	IPR002521; HCV_core.			
DR	InterPro;	IPR002519; HCV env.			
DR	InterPro;	IPR002531; HCV NS1.			
DR	InterPro;	IPR002518; HCV NS2.			
DR	InterPro;	IPR000745; HCV NS4a.			
DR	InterPro;	IPR001490; HCV NS4b.			
DR	InterPro;	IPR002868; HCV NS5a.			
DR	InterPro;	IPR002166; HCV RdRp.			
DR	InterPro;	IPR001650; Helicase_C.			
DR	InterPro;	IPR007095; RNA pol DS_PS.			
DR	InterPro;	IPR007094; RNA pol_PSVlr.			
DR	Pfam;	PF01543; HCV_capsid; 1.			
DR	Pfam;	PF01542; HCV_core; 1.			
DR	Pfam;	PF01539; HCV env; 1.			
DR	Pfam;	PF01560; HCV NS1; 1.			
DR	Pfam;	PF01538; HCV NS2; 1.			
DR	Pfam;	PF02907; HCV NS3; 1.			
DR	Pfam;	PF01006; HCV NS4a; 1.			
DR	Pfam;	PF01001; HCV NS4b; 1.			
DR	Pfam;	PF01506; HCV NS5a; 1.			
DR	Pfam;	PF00271; helicase_C; 1.			
DR	Pfam;	PF00998; viral_RdRp; 1.			
DR	ProDom;	PD185062; HCV NS1; 1.			
DR	SMART;	SM00487; DEXdc; 1.			
DR	PROSITE;	PS00190; CYTOCHROME C; 1.			
CA		Coat protein, Envelope protein, Glycoprotein, Nonstructural protein;			
FM		Polyprotein, RNA-directed RNA polymerase, Transferase, Transmembrane			

SQ SEQUENCE 3010 AA; 327234 MW; 44C34677649CE8DD CRC64;  
 Query Match 86.7%; Score 1535; DB 12; Length 3010;  
 Best Local Similarity 94.1%; Pred. No. 1.2e-124;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFPAAGLIRACMLVRKAAGHYQVAFMLALALGTYYVDHLTPLODVAHAG 75  
 DB 904 AGITRVYFPAAGLIRACMLVRKAAGHYQVAFMLALALGTYYVDHLTPLODVAHAG 963  
 QY 76 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135  
 DB 964 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 1023  
 QY 136 RLAPITAYSCQTRGLIGCIITSLTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVYH 195  
 DB 1024 RLAPITAYSCQTRGLIGCIITSLTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVYH 1083  
 QY 196 GAGSKTLAAGPKPITOMYTNVODLVGMQAPPGARSMTPTCGSSDLYLVRHADYIPVR 255  
 DB 1084 GAGSKTLAAGPKPITOMYTNVODLVGMQAPPGARSMTPTCGSSDLYLVRHADYIPVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 14  
 Q9J3H2 PRELIMINARY; PRT; 3010 AA.  
 ID Q9J3H2  
 AC Q9J3H2  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 OX NCBI\_Taxid=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MM20;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NCV-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL, AF207761; AAE65951.1; -.  
 DR PIR, A61196; A61196.  
 DR PIR, P00246; P00246.  
 DR PIR, P50329; P50329.  
 DR HSP, P26663; INS3.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transport activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003668; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0006236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; P: electron transport; IEA.  
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.

DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_typsin.  
 DR InterPro: IPR000345; CysC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002568; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART, SM00487; DEXDC; 1.  
 DR PROSITE, PS00190; CYTOCHROME C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326763 MW; 1A4BBE4BE51440D0 CRC64;

Query Match 86.6%; Score 1534; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 1.4e-124;  
 Matches 288; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFPAAGLIRACMLVRKAAGHYQVAFMLALALGTYYVDHLTPLODVAHAG 75  
 DB 904 AGITRVYFPAAGLIRACMLVRKAAGHYQVAFMLALALGTYYVDHLTPLODVAHAG 963  
 QY 76 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 135  
 DB 964 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGM 1023  
 QY 136 RLAPITAYSCQTRGLIGCIITSLTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVYH 195  
 DB 1024 RLAPITAYSCQTRGLIGCIITSLTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVYH 1083  
 QY 196 GAGSKTLAAGPKPITOMYTNVODLVGMQAPPGARSMTPTCGSSDLYLVRHADYIPVR 255  
 DB 1084 GAGSKTLAAGPKPITOMYTNVODLVGMQAPPGARSMTPTCGSSDLYLVRHADYIPVR 1143  
 QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 316 TMRT 319  
 DB 1204 TMRS 1207

RESULT 15  
 Q9J3I0 PRELIMINARY; PRT; 3010 AA.  
 ID Q9J3I0  
 AC Q9J3I0  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Genome polypeptide.

OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD12;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DDA databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AF207753; AAF55943.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; LUXP.  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F: electron transporter activity; IEA.  
 DR GO; GO:0016787; F: hydrolase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006118; F: electron transport; IEA.  
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; F: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral genome replication; IEA.  
 DR InterPro: IPR003003; Cys Ser trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD16062; HCV\_NS1; 1.  
 DR SMART; SMO0487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 KM ATP-binding, Coat protein, Envelope protein, Glycoprotein, Helicase;  
 KM Hydrolyase, Nonstructural protein, Polypeptide, Polypeptide,  
 KM RNA-directed RNA polymerase, Transferase, Transmembrane.  
 SQ SEQUENCE 3010 AA; 32692 MW; 074098DB305AF1A9 CRC64;

Query Match 66.6%; Score 1534; DB 12; Length 3010;

Best Local Similarity 95.1%; Pred. No. 1.4e-124;  
 Matches 289; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
 QY 16 AGITKVPYFVAAOGLIRACMLVRKAGGHVQVAFMKLAALGTGYVDHLTPLODMAHAG 75  
 DB 904 AGITRVPYFVAAOGLIRACMLVRKAGGHVQVAFMKLAALGTGYVDHLTPLRGMAHTG 963  
 QY 76 LRDLAVAEPIFSDMEKXITTWGADTPAACDIIISGLPVASRRREIILGPADDFEGQW 135  
 DB 964 LRDLAVAEPIFSDMEKXITTWGADTPAACDIIISGLPVASRRREIILGPADDFEGQW 1023  
 QY 136 RLAPITVASSQOTRGLLCITTSITGRDKQVEGEVQVSTATQSFPLATCVNGVCMVTFH 195  
 DB 1024 RLAPITVASSQOTRGLLCITTSITGRDKQVEGEVQVSTATQSFPLATCVNGVCMVTFH 1083  
 QY 196 GAGSKTLAEPKPIPTOMTNTDQVNGQAPPGARSMTPCTCGSSDLYLTRADVTPVR 255  
 DB 1084 GAGSKTLAEPKPIPTOMTNTDQVNGQAPPGARSMTPCTCGSSDLYLTRADVTPVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHVGIFRAAVCTGKAVADFIPESEMET 315  
 DB 1144 RRGDSRGSILSPRPVSYLKSSGGPILCPSGHVGIFRAAVCTGKAVADFIPESEMET 1203  
 QY 316 TWRT 319  
 DB 1204 TWRS 1207

Search completed: May 6, 2004, 09:35:44  
 Job time : 34.2384 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.109 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKKLEHHHHHTSAGITK.....TTMKTSAMRHPQGGKKKK 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUTS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1531	86.4	2201	4	US-09-539-601-6
2	1531	86.4	2201	4	US-09-539-601-15
3	1531	86.4	3010	4	US-09-539-601-3
4	1531	86.4	3010	4	US-09-539-601-21
5	1531	86.4	3010	4	US-09-539-601-27
6	1528	86.3	1692	4	US-09-263-933-4
7	1528	86.3	1692	4	US-09-919-901-4
8	1528	86.3	2307	3	US-09-263-933-2
9	1528	86.3	2307	3	US-09-919-901-2
10	1525	86.1	1692	4	US-09-263-933-11
11	1525	86.1	1692	4	US-09-919-901-11
12	1525	86.1	2307	3	US-09-263-933-9
13	1525	86.1	2307	3	US-09-919-901-9
14	1524	86.1	3010	4	US-09-539-601-33
15	1516	85.6	1692	3	US-09-263-933-18
16	1516	85.6	1692	3	US-09-919-901-18
17	1516	85.6	2307	4	US-09-263-933-16
18	1516	85.6	2307	4	US-09-919-901-16
19	1505	85.0	3010	3	US-09-014-416-3
20	1479	83.5	2013	1	US-08-324-977-12
21	1479	83.5	2013	2	US-08-364-616-12
22	1479	83.5	2013	2	US-08-904-686A-12
23	1479	83.5	2013	2	US-09-315-850-12
24	1479	83.5	2201	4	US-08-952-961A-2
25	1479	83.5	2620	1	US-08-324-977-32
26	1479	83.5	2620	2	US-08-364-616-32
27	1479	83.5	2620	2	US-08-904-686A-32

28	1479	83.5	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1479	83.5	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1479	83.5	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1479	83.5	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1479	83.5	2621	2	US-09-315-850-36	Sequence 36, Appl
33	1479	83.5	3010	1	US-08-324-977-2	Sequence 2, Appl
34	1479	83.5	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1479	83.5	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1479	83.5	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1479	83.5	3010	2	US-08-904-686A-2	Sequence 2, Appl
38	1479	83.5	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1479	83.5	3010	3	US-09-315-850-2	Sequence 2, Appl
40	1479	83.5	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1406	79.4	3012	3	US-08-811-566-2	Sequence 2, Appl
42	1406	79.4	3012	4	US-09-034-756-2	Sequence 2, Appl
43	1405	79.3	2894	4	US-08-466-975A-23	Sequence 23, Appl
44	1405	79.3	2894	2	US-08-391-671A-23	Sequence 23, Appl
45	1405	79.3	2894	3	US-08-467-902A-23	Sequence 23, Appl

## ALIGNMENTS

RESULT 1	US-09-539-601-6	US-09-539-601-6
;	Sequence 6, Application US/09539601C	
;	Patent No. 6638343	
;	GENERAL INFORMATION:	
;	APPLICANT: Bartschlagel, Ralf FW	
;	TITLE OF INVENTION: Hepatitis C Virus Cell Culture System	
;	FILE REFERENCE: all sequences	
;	CURRENT APPLICATION NUMBER: US/09/539,601C	
;	CURRENT FILING DATE: 2001-08-30	
;	EARLIER APPLICATION NUMBER: 199-15 178.4 GERMANY	
;	EARLIER FILING DATE: 1999-04-03	
;	NUMBER OF SEQ ID NOS: 51	
;	SOFTWARE: Patentin Ver. 2.1	
;	SEQ ID NO 6	
;	LENGTH: 2201	
;	TYPE: PRT	
;	ORGANISM: Hepatitis C virus	
;	US-09-539-601-6	

Query Match	86.4%; Score 1531; DB 4; Length 2201;
Best Local Similarity	94.7%; Pred. No. 26-143; 7; Indels 0; Gaps 0;
Matches	288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY	16 AGITKVPYVRAQGLIRACMLVYRKAAGHYVOMAFKLAALTGTYYDHLTPLQDMANAG	75
DB	95 AGITKVPYVRAHGLIRACMLVYRKAAGHYVOMAFKLAALTGTYYDHLTPLQDMANAG	154
QY	76 LRDLAAVPEPVIFSDMEVKIITMGADTAACGDIISGLPVASARGREIILGPADNFEQGM	135
DB	155 LRDLAAVPEPVIFSDMEVKIITMGADTAACGDIISGLPVASARGREIILGPADNFEQGM	214
QY	136 RLIAPIYVSOQTRGLIGCIITSLGRDNQYEGEVQVSTATQSLFATCVGVCWTVH	195
DB	215 RLIAPIYVSOQTRGLIGCIITSLGRDNQYEGEVQVSTATQSLFATCVGVCWTVH	274
QY	196 GAGSKTLAPKPGITMTMNVODLVGKQAPPGASMPCTGSSSLYVTHADVIPIR	255
DB	275 GAGSKTLAPKPGITMTMNVODLVGKQAPPGASMPCTGSSSLYVTHADVIPIR	334
QY	256 RRDSSGSLISRPVSYLKGSSGGLLCPGSAVGIFFRAAVCTRGVAKAVDEIPVESMET	315
DB	335 RRDSSGSLISRPVSYLKGSSGGLLCPGSAVGIFFRAAVCTRGVAKAVDEIPVESMET	394
QY	316 TMTT 319	
DB	395 TMTT 398	

RESULT 2

US-09-539-601-15  
; Sequence 15, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-15

Query Match 86.4%; Score 1531; DB 4; Length 2201;  
Best Local Similarity 94.7%; Pred. No. 2e-143;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVFYFRAQGLIRACMLVRKAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAG 75  
DB 95 AGITKVFYFRAHGLIRACMLVRKAGHYVQMALMKLAALTGYVVDHLTPLODMAHAG 154  
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135  
DB 155 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 214  
QY 136 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVYH 195  
DB 215 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVYH 274  
QY 196 GAGSKTLGPKGPITQMTYNDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 255  
DB 275 GAGSKTLGPKGPITQMTYNDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 334  
QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 315  
DB 335 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 394  
QY 316 TMRT 319  
DB 395 TMRG 396

RESULT 3  
US-09-539-601-3  
; Sequence 3, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-3

Query Match 86.4%; Score 1531; DB 4; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 3.2e-143;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVFYFRAQGLIRACMLVRKAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAG 75

DB 904 AGITKVFYFRAHGLIRACMLVRKAGHYVQMALMKLAALTGYVVDHLTPLODMAHAG 963  
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135  
DB 964 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 1023  
QY 136 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVYH 195  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVYH 1083  
QY 196 GAGSKTLGPKGPITQMTYNDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 255  
DB 1084 GAGSKTLGPKGPITQMTYNDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 1143  
QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 315  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 1203  
QY 316 TMRT 319  
DB 1204 TMRG 1207

RESULT 4  
US-09-539-601-21  
; Sequence 21, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-21

Query Match 86.4%; Score 1531; DB 4; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 3.2e-143;  
Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVFYFRAQGLIRACMLVRKAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAG 75  
DB 904 AGITKVFYFRAHGLIRACMLVRKAGHYVQMALMKLAALTGYVVDHLTPLODMAHAG 963  
QY 76 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135  
DB 964 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 1023  
QY 136 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVYH 195  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCNGVCMTVYH 1083  
QY 196 GAGSKTLGPKGPITQMTYNDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 255  
DB 1084 GAGSKTLGPKGPITQMTYNDODLVGQAPPGARSMTPTCGSSDLYLTRHADVI PVR 1143  
QY 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 315  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVFEVESMET 1203  
QY 316 TMRT 319  
DB 1204 TMRG 1207

RESULT 5  
 US-09-539-601-27  
 ; Sequence 27, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlagier, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 27  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-27

Query Match 86.4%; Score 1531; DB 4; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 3.2e-143;  
 Matches 288; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 75  
 |||||  
 DB 904 AGITKVPYFVRAHGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 963  
 QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 135  
 |||||  
 DB 964 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 1023  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFH 195  
 |||||  
 DB 1024 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFH 1083  
 QY 196 GAGSKTLGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255  
 |||||  
 DB 1084 GAGSKTLGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 1143  
 QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPGSHANGITRAVCTRGVAKADPIPVESMET 315  
 |||||  
 DB 1144 RRGDSRGSILSPRPVSYLKSSGGPILCPGSHANGITRAVCTRGVAKADPIPVESMET 1203  
 QY 316 TMRAT 319  
 |||||  
 DB 1204 TMRAT 1207

RESULT 6  
 US-09-263-933-4  
 ; Sequence 4, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/263,933  
 ; CURRENT FILING DATE: 1999-03-08  
 ; EARLIER APPLICATION NUMBER: 09/129,611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 US-09-263-933-4  
 Query Match 86.3%; Score 1528; DB 3; Length 1692;

Best Local Similarity 94.1%; Pred. No. 2.7e-143;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 75  
 |||||  
 DB 183 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 242  
 QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 135  
 |||||  
 DB 243 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFH 195  
 |||||  
 DB 303 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFH 362  
 QY 196 GAGSKTLGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255  
 |||||  
 DB 363 GAGSKTLGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 422  
 QY 256 RRGDSRGSILSPRPVSYLKSSGGPILCPGSHANGITRAVCTRGVAKADPIPVESMET 315  
 |||||  
 DB 423 RRGDSRGSILSPRPVSYLKSSGGPILCPGSHANGITRAVCTRGVAKADPIPVESMET 482  
 QY 316 TMRAT 319  
 |||||  
 DB 483 TMRAT 486

RESULT 7  
 US-09-919-901-4  
 ; Sequence 4, Application US/09919901  
 ; Patent No. 6599738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 US-09-919-901-4

Query Match 86.3%; Score 1528; DB 4; Length 1692;  
 Best Local Similarity 94.1%; Pred. No. 2.7e-143;  
 Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 75  
 |||||  
 DB 183 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLODMAHAG 242  
 QY 76 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 135  
 |||||  
 DB 243 LRDIAVAEVPVFSMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
 QY 136 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFH 195  
 |||||  
 DB 303 RLAPITAYSOOTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFH 362  
 QY 196 GAGSKTLGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255  
 |||||

Db 363 GAGSKTLAAGPKGPIQMTNTVDODLVGMQAPPGARSITPCTCGSSDLVLTTRHADV1PVR 422  
QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
Db 423 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 482  
QY 316 TMRT 319  
Db 483 TMRS 486

RESULT 8  
US-09-263-933-2  
Sequence 2, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PR  
ORGANISM: Artificial Sequence  
US-09-263-933-2

Query Match 86.3%; Score 1528; DB 3; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 4.3e-143;  
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLALITGYVVDHLTPLODMAHAG 75  
Db 275 AGITRVYFVRAOGLIRACMLVRKAAGHYVQAFMKLALITGYVVDHLTPLODMAHAG 334  
QY 76 LRDVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
Db 335 LRDVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 394  
QY 136 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 195  
Db 395 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 454  
QY 196 GAGSKTLAAGPKGPIQMTNTVDODLVGMQAPPGARSMTCTCGSSDLVLTTRHADV1PVR 255  
Db 455 GAGSKTLAAGPKGPIQMTNTVDODLVGMQAPPGARSMTCTCGSSDLVLTTRHADV1PVR 514  
QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
Db 515 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 574  
QY 316 TMRT 319  
Db 575 TMRS 578

RESULT 9  
US-09-919-901-2  
Sequence 2, Application US/09919901  
Patent No. 6599738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PR  
ORGANISM: Artificial Sequence  
US-09-919-901-2

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PR  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-2

Query Match 86.3%; Score 1528; DB 4; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 4.3e-143;  
Matches 286; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLALITGYVVDHLTPLODMAHAG 75  
Db 275 AGITRVYFVRAOGLIRACMLVRKAAGHYVQAFMKLALITGYVVDHLTPLODMAHAG 334  
QY 76 LRDVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
Db 335 LRDVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 394  
QY 136 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 195  
Db 395 RLAPITAYSOQTRGLICITISLTGRDKNOVEGEVQVSTATOSFLATCNVGCMTVPH 454  
QY 196 GAGSKTLAAGPKGPIQMTNTVDODLVGMQAPPGARSMTCTCGSSDLVLTTRHADV1PVR 255  
Db 455 GAGSKTLAAGPKGPIQMTNTVDODLVGMQAPPGARSMTCTCGSSDLVLTTRHADV1PVR 514  
QY 256 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
Db 515 RRGDSRGSLLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 574  
QY 316 TMRT 319  
Db 575 TMRS 578

RESULT 10  
US-09-263-933-11  
Sequence 11, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PR  
ORGANISM: Artificial Sequence  
US-09-263-933-11

Query Match 86.1%; Score 1525; DB 3; Length 1692;  
Best Local Similarity 93.8%; Pred. No. 5.4e-143;  
Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;



```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYVDHLTPLODMAHAG 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYVDHLTPLODMAHAG 242
QY 76 LRDIAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNPEGQGW 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 LRDIAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNPEGQGW 302
QY 136 RLAPITAYSQOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 RLAPITAYSQOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVYH 362
QY 196 GAGSKTLAAGPKPIITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 GAGSKTLAAGPKPIITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 422
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIVESMET 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 423 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIVESMET 482
QY 316 TMR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 483 TMR 486

```

```

RESULT 11
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6593738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-11

```

Query Match 86.1%; Score 1525; DB 4; Length 1692;  
 Best Local Similarity 93.8%; Pred. No. 5.4e-143;  
 Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYVDHLTPLODMAHAG 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYVDHLTPLODMAHAG 242
QY 76 LRDIAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNPEGQGW 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 LRDIAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNPEGQGW 302
QY 136 RLAPITAYSQOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 RLAPITAYSQOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVYH 362
QY 196 GAGSKTLAAGPKPIITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 GAGSKTLAAGPKPIITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 422
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIVESMET 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 423 RRGDSRGSLLSPRPVSYLKSGAGGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIVESMET 482
QY 316 TMR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 483 TMR 486

```

```

RESULT 12
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

```

Query Match 86.1%; Score 1525; DB 3; Length 2307;  
 Best Local Similarity 93.8%; Pred. No. 8.5e-143;  
 Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYVDHLTPLODMAHAG 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGYTYVDHLTPLODMAHAG 334
QY 76 LRDIAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNPEGQGW 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 335 LRDIAVAVEPVYFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNPEGQGW 394
QY 136 RLAPITAYSQOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVYH 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 395 RLAPITAYSQOQTRGLIGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVYH 454
QY 196 GAGSKTLAAGPKPIITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 455 GAGSKTLAAGPKPIITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 514
QY 256 RRGDSRGSLLSPRPVSYLKSGSGGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIVESMET 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 515 RRGDSRGSLLSPRPVSYLKSGAGGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIVESMET 574
QY 316 TMR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 575 TMR 578

```

```

RESULT 13
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6593738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02

```

PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-9

Query Match 86.1%; Score 1525; DB 4; Length 2307;  
Best Local Similarity 93.8%; Pred. No. 8.5e-143;  
Matches 285; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
DB 275 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 334  
QY 76 LRDLAVALVEPVFSDMETKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
DB 335 LRDLAVALVEPVFSDMETKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 394  
QY 136 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTAQSFLATCVNGVCTVYH 195  
DB 395 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTAQSFLATCVNGVCTVYH 454  
QY 196 GAGSKTLAAGPKPITOMTNTVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIYVR 255  
DB 455 GAGSKTLAAGPKPITOMTNTVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIYVR 514  
QY 256 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315  
DB 515 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 574  
QY 316 TMR 319  
DB 575 TMR 578

RESULT 14  
US-09-539-601-33  
Sequence 33, Application US/09539601C

Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartschlag, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 33  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C Virus  
US-09-539-601-33

Query Match 86.1%; Score 1524; DB 4; Length 3010;  
Best Local Similarity 94.4%; Pred. No. 1.6e-142;  
Matches 287; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
DB 904 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
QY 76 LRDLAVALVEPVFSDMETKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135

DB 964 LRDLAVALVEPVFSDMETKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 1023  
QY 136 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTAQSFLATCVNGVCTVYH 195  
DB 1024 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTAQSFLATCVNGVCTVYH 1083  
QY 196 GAGSKTLAAGPKPITOMTNTVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIYVR 255  
DB 1084 GAGSKTLAAGPKPITOMTNTVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIYVR 1143  
QY 256 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315  
DB 1144 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
QY 316 TMR 319  
DB 1204 TMR 1207

RESULT 15  
US-09-263-933-18  
Sequence 18, Application US/09263933

Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0003A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 18  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-18

Query Match 85.6%; Score 1516; DB 3; Length 1692;  
Best Local Similarity 93.4%; Pred. No. 4.3e-142;  
Matches 284; Conservative 13; Mismatches 7; Indels 0; Gaps 0;

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
DB 183 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 242  
QY 76 LRDLAVALVEPVFSDMETKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135  
DB 243 LRDLAVALVEPVFSDMETKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 302  
QY 136 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTAQSFLATCVNGVCTVYH 195  
DB 303 RLAPITAYSQOTRGLGCIITSLTGDKNQVEGEVYSTAQSFLATCVNGVCTVYH 362  
QY 196 GAGSKTLAAGPKPITOMTNTVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIYVR 255  
DB 363 GAGSKTLAAGPKPITOMTNTVDOLVGMQAPPGARSMTPTCCSSDLVLTTRHADVIYVR 422  
QY 256 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315  
DB 423 RRGDSRGSLSPPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 482  
QY 316 TMR 319  
DB 483 TMR 486

Search completed: May 6, 2004, 09:39:00  
Job time: 14.109 secs

Fri May 7 13:37:08 2004

us-10-650-585-4.rai

---

Page 7



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 34.167 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-4  
Perfect score: 1771  
Sequence: 1 MKKKLEHHHHHTSAGITK.....TTMTSSAMRHPRGKXXX 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1771	100.0	334	US-10-017-736-4
2	1771	100.0	334	US-10-650-585-4
3	1660	93.7	409	US-10-017-736-2
4	1660	93.7	409	US-10-650-585-2
5	1589	89.7	303	US-10-017-736-10
6	1589	89.7	303	US-10-650-585-10
7	1589	89.7	341	US-10-017-736-14
8	1589	89.7	341	US-10-650-585-14
9	1589	89.7	352	US-10-017-736-13
10	1589	89.7	352	US-10-650-585-13
11	1589	89.7	380	US-10-017-736-12
12	1589	89.7	380	US-10-650-585-12
13	1589	89.7	393	US-10-017-736-11
14	1589	89.7	393	US-10-650-585-11
15	1580	89.2	303	US-10-017-736-18

16	1580	89.2	303	US-10-650-585-18	Sequence 18, Appl
17	1579	89.2	303	US-10-017-736-16	Sequence 16, Appl
18	1579	89.2	303	US-10-650-585-16	Sequence 16, Appl
19	1570	88.7	301	US-10-017-736-17	Sequence 17, Appl
20	1570	88.7	301	US-10-650-585-17	Sequence 17, Appl
21	1532	86.5	222	US-10-017-736-15	Sequence 15, Appl
22	1532	86.5	222	US-10-650-585-15	Sequence 15, Appl
23	1531	86.4	220	US-10-029-907-3	Sequence 3, Appl
24	1531	86.4	220	US-10-309-561-3	Sequence 3, Appl
25	1531	86.4	3010	US-10-467-000-1	Sequence 1, Appl
26	1528	86.3	1592	US-09-919-901-4	Sequence 4, Appl
27	1528	86.3	1592	US-10-191-966-4	Sequence 4, Appl
28	1528	86.3	2307	US-09-919-901-2	Sequence 2, Appl
29	1528	86.3	2307	US-10-191-966-2	Sequence 2, Appl
30	1525	86.1	1692	US-09-919-901-11	Sequence 11, Appl
31	1525	86.1	1692	US-10-191-966-11	Sequence 11, Appl
32	1525	86.1	2307	US-09-919-901-9	Sequence 9, Appl
33	1525	86.1	2307	US-10-191-966-9	Sequence 9, Appl
34	1516	85.6	1692	US-10-191-966-18	Sequence 18, Appl
35	1516	85.6	1692	US-10-191-966-18	Sequence 18, Appl
36	1516	85.6	2307	US-09-919-901-16	Sequence 16, Appl
37	1516	85.6	2307	US-10-191-966-16	Sequence 16, Appl
38	1479	83.5	2201	US-10-085-476-2	Sequence 2, Appl
39	1479	83.5	2201	US-09-742-659-4	Sequence 4, Appl
40	1406	79.4	3011	US-09-891-894-3	Sequence 3, Appl
41	1406	79.4	3011	US-10-184-150-3	Sequence 3, Appl
42	1406	79.4	3011	US-10-328-997-3	Sequence 3, Appl
43	1406	79.4	3012	US-09-238-076-2	Sequence 2, Appl
44	1406	79.4	3012	US-09-995-937-2	Sequence 2, Appl
45	1406	79.4	3012	US-09-917-563-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-10-017-736-4  
; Sequence 4, Application US/10017736  
; Publication NO: US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 334  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-4

Query Match	100.0%	Score 1771	DB 13	Length 334
Best Local Similarity	100.0%	Pred. No. 8.2e+163		
Matches 334	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	1	1	1
DB	1	1	1	1
QY	61	61	61	61
DB	61	61	61	61
QY	121	121	121	121
DB	121	121	121	121
QY	181	181	181	181
DB	181	181	181	181

Db 181 FLATCVNGVCMVTFHAGSGKTLGPKGPIQTQMTTNDQDVLVGMQAPPGASMTPTCTGSS 240  
 QY 241 DLYVTRHADVIPIVRRRSGDSRGLSPVSYLKSGSGGGLCPSGHAGIFRAAVCTRG 300  
 Db 241 DLYVTRHADVIPIVRRRSGDSRGLSPVSYLKSGSGGGLCPSGHAGIFRAAVCTRG 300  
 QY 301 VAKAVDFIPVESMETTMTTSAMRHPOFGKXXX 334  
 Db 301 VAKAVDFIPVESMETTMTTSAMRHPOFGKXXX 334

## RESULT 2

US-10-650-585-4  
 ; Sequence 4, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 334  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-4

Query Match 100.0%; Score 1771; DB 16; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-163;  
 Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K K K K L E H H H H H T S A G I T K V P F V R A G L I R A C M L V R K A G H Y V M A F M K L A A L T G T Y 60  
 Db 1 M K K K L E H H H H H T S A G I T K V P F V R A G L I R A C M L V R K A G H Y V M A F M K L A A L T G T Y 60  
 QY 61 V Y D H L T P L Q D W A H A G L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R 120  
 Db 61 V Y D H L T P L Q D W A H A G L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R 120  
 QY 121 E I L L G P A D N F E G Q G R L A P I T A Y S Q Q T R G L G C I I T S L T R D K N O V E G E V Q V S T A T Q S 180  
 Db 121 E I L L G P A D N F E G Q G R L A P I T A Y S Q Q T R G L G C I I T S L T R D K N O V E G E V Q V S T A T Q S 180  
 QY 181 F L A T C V N G V C M T V F H A G S K T L A G K G P I T Q M T T N D Q D L V G M Q A P P G A S M T P C T G S S 240  
 Db 181 F L A T C V N G V C M T V F H A G S K T L A G K G P I T Q M T T N D Q D L V G M Q A P P G A S M T P C T G S S 240  
 QY 241 D L Y V T R H A D V I P V R R R S G D S R G L S P R V S Y L K S G S G G L C P S G H A G I F R A A V C T R G 300  
 Db 241 D L Y V T R H A D V I P V R R R S G D S R G L S P R V S Y L K S G S G G L C P S G H A G I F R A A V C T R G 300  
 QY 301 V A K A V D F I P V E S M E T T M T S A M R H P O F G K K X X 334  
 Db 301 V A K A V D F I P V E S M E T T M T S A M R H P O F G K K X X 334

## RESULT 3

US-10-017-736-2  
 ; Sequence 2, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 409  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-017-736-2

Query Match 93.7%; Score 1660; DB 13; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 6e-152;  
 Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 A G I T K V P F V R A G L I R A C M L V R K A G H Y V M A F M K L A A L T G T Y V Y D H L T P L Q D W A H A G 75  
 Db 95 A G I T K V P F V R A G L I R A C M L V R K A G H Y V M A F M K L A A L T G T Y V Y D H L T P L Q D W A H A G 154  
 QY 76 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R E I L L G P A D N F E G Q G W 135  
 Db 155 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R E I L L G P A D N F E G Q G W 214  
 QY 136 R L L A P I T A Y S Q Q T R G L G C I I T S L T R D K N O V E G E V Q V S T A T Q S F L A T C V N G V C M T V F H 195  
 Db 215 R L L A P I T A Y S Q Q T R G L G C I I T S L T R D K N O V E G E V Q V S T A T Q S F L A T C V N G V C M T V F H 274  
 QY 196 G A G S K T L A G K G P I T Q M T T N D Q D L V G M Q A P P G A S M T P C T G S S D L Y L T R H A D V I P V R 255  
 Db 275 G A G S K T L A G K G P I T Q M T T N D Q D L V G M Q A P P G A S M T P C T G S S D L Y L T R H A D V I P V R 334  
 QY 256 R R G D S R G S I L S P R P V S Y L K S G S G G L C P S G H A G I F R A A V C T G V A K A V D F I P V E S M E T 315  
 Db 335 R R G D S R G S I L S P R P V S Y L K S G S G G L C P S G H A G I F R A A V C T G V A K A V D F I P V E S M E T 394  
 QY 316 T W R T S S A M R H P O F G K 330  
 Db 395 T W R T S S A M R H P O F G K 409

## RESULT 4

US-10-650-585-2  
 ; Sequence 2, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 409  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-2

Query Match 93.7%; Score 1660; DB 16; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 6e-152;  
 Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 A G I T K V P F V R A G L I R A C M L V R K A G H Y V M A F M K L A A L T G T Y V Y D H L T P L Q D W A H A G 75  
 Db 95 A G I T K V P F V R A G L I R A C M L V R K A G H Y V M A F M K L A A L T G T Y V Y D H L T P L Q D W A H A G 154  
 QY 76 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R E I L L G P A D N F E G Q G W 135  
 Db 155 L R D L A V A V E P V I F S D M E V K I I T W G A D T A A C G D I I S G L P V S A R G R E I L L G P A D N F E G Q G W 214  
 QY 136 R L L A P I T A Y S Q Q T R G L G C I I T S L T R D K N O V E G E V Q V S T A T Q S F L A T C V N G V C M T V F H 195

```

Db 215 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCNGVCWTFVH 274
QY 196 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSTPTCTGSSDLYLVTREADVIVR 255
Db 275 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSTPTCTGSSDLYLVTREADVIVR 334
QY 256 RRGDSRGSLLSPREVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 335 RRGDSRGSLLSPREVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 394
QY 316 TMRSSAMRHPOFGG 330
Db 395 TMRSSAMRHPOFGG 409

```

RESULT 5

```

US-10-017-736-10
; Sequence 10, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-10

```

Query Match 89.7%; Score 1589; DB 13; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 3e-145;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135
Db 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 120
QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCNGVCWTFVH 195
Db 121 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCNGVCWTFVH 180
QY 196 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSTPTCTGSSDLYLVTREADVIVR 255
Db 181 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSTPTCTGSSDLYLVTREADVIVR 240
QY 256 RRGDSRGSLLSPREVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 241 RRGDSRGSLLSPREVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
QY 316 TMR 318
Db 301 TMR 303

```

RESULT 6

```

US-10-650-585-10
; Sequence 10, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082

```

```

; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-10

```

Query Match 89.7%; Score 1589; DB 16; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 3e-145;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135
Db 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 120
QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCNGVCWTFVH 195
Db 121 RLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCNGVCWTFVH 180
QY 196 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSTPTCTGSSDLYLVTREADVIVR 255
Db 181 GAGSKTLAAGPKPITOMYTNVDOLVGMQAPPGARSTPTCTGSSDLYLVTREADVIVR 240
QY 256 RRGDSRGSLLSPREVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 315
Db 241 RRGDSRGSLLSPREVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
QY 316 TMR 318
Db 301 TMR 303

```

RESULT 7

```

US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-14

```

Query Match 89.7%; Score 1589; DB 13; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-145;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
Db 39 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
QY 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGM 135

```

```

Db      99 LRLDAVAEPIVFSDMVEKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 158
;
QY      136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLFATCVNGVCTVPH 195
;
Db      159 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLFATCVNGVCTVPH 218
;
QY      196 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIVR 255
;
Db      219 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIVR 278
;
QY      256 RRGDSRGSILSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
;
Db      279 RRGDSRGSILSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338
;
QY      316 TMR 318
;
Db      339 TMR 341

```

```

RESULT 8
US-10-650-585-14
; Sequence 14, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-14

```

```

Query Match      89.7%; Score 1589; DB 16; Length 341;
Best Local Similarity 100.0%; Pred. No. 3.5e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      16 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
;
Db      39 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
;
QY      76 LRLDAVAEPIVFSDMVEKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135
;
Db      99 LRLDAVAEPIVFSDMVEKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 158
;
QY      136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLFATCVNGVCTVPH 195
;
Db      159 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLFATCVNGVCTVPH 218
;
QY      196 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIVR 255
;
Db      219 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIVR 278
;
QY      256 RRGDSRGSILSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
;
Db      279 RRGDSRGSILSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338
;
QY      316 TMR 318
;
Db      339 TMR 341

```

```

RESULT 9
US-10-017-736-13
; Sequence 13, Application US/10017736

```

```

; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 352
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-13

```

```

Query Match      89.7%; Score 1589; DB 13; Length 352;
Best Local Similarity 100.0%; Pred. No. 3.7e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      16 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 75
;
Db      50 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 109
;
QY      76 LRLDAVAEPIVFSDMVEKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 135
;
Db      110 LRLDAVAEPIVFSDMVEKIIITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGW 169
;
QY      136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLFATCVNGVCTVPH 195
;
Db      170 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSLFATCVNGVCTVPH 229
;
QY      196 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIVR 255
;
Db      230 GAGSKTLAGPKGPITQMTYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIVR 289
;
QY      256 RRGDSRGSILSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
;
Db      290 RRGDSRGSILSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 349
;
QY      316 TMR 318
;
Db      350 TMR 352

```

```

RESULT 10
US-10-650-585-13
; Sequence 13, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 352
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-13

```

```

Query Match      89.7%; Score 1589; DB 16; Length 352;
Best Local Similarity 100.0%; Pred. No. 3.7e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      16 AGITKVPYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMAHAG 75

```



```

Db 50 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAG 109
Qy 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135
Db 110 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 169
Qy 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 195
Db 170 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 229
Qy 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVI 255
Db 230 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVI 289
Qy 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315
Db 290 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 349
Qy 316 TMR 318
Db 350 TMR 352

```

```

RESULT 11
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12

```

```

Query Match 89.7%; Score 1589; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 4,1e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAG 75
Db 78 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAG 137
Qy 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135
Db 138 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 197
Qy 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 195
Db 198 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 257
Qy 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVI 255
Db 258 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVI 317
Qy 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315
Db 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 377
Qy 316 TMR 318
Db 378 TMR 380

```

```

RESULT 12
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-12

```

```

Query Match 89.7%; Score 1589; DB 16; Length 380;
Best Local Similarity 100.0%; Pred. No. 4,1e-145;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 16 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAG 75
Db 78 AGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAG 137
Qy 76 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135
Db 138 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 197
Qy 136 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 195
Db 198 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 257
Qy 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVI 255
Db 258 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVI 317
Qy 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 315
Db 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 377
Qy 316 TMR 318
Db 378 TMR 380

```

```

RESULT 13
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

```

```

Query Match 89.7%; Score 1589; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 4,3e-145;

```

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGYVVDHLLTPLODMAHAG 75
DB 91 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGYVVDHLLTPLODMAHAG 150
QY 76 LRDLAAVEVPYFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135
DB 151 LRDLAAVEVPYFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 210
QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 195
DB 211 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 270
QY 196 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255
DB 271 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 330
QY 256 RRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
DB 331 RRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 390
QY 316 TMR 318
DB 391 TMR 393

```

## RESULT 14

```

US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

```

Query Match 89.7%; Score 1589; DB 16; Length 393;

Best Local Similarity 100.0%; Pred. No. 4,3e-145; Mismatches 0; Indels 0; Gaps 0;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGYVVDHLLTPLODMAHAG 75
DB 91 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGYVVDHLLTPLODMAHAG 150
QY 76 LRDLAAVEVPYFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135
DB 151 LRDLAAVEVPYFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 210
QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 195
DB 211 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 270
QY 196 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255
DB 271 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 330
QY 256 RRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
DB 331 RRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 390

```

QY 316 TMR 318  
DB 391 TMR 393

## RESULT 15

```

US-10-017-736-18
; Sequence 18, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 303
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-18

```

Query Match 89.2%; Score 1580; DB 13; Length 303;

Best Local Similarity 99.7%; Pred. No. 2,2e-144; Mismatches 1; Indels 0; Gaps 0;

Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 16 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGYVVDHLLTPLODMAHAG 75
DB 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGYVVDHLLTPLODMAHAG 60
QY 76 LRDLAAVEVPYFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 135
DB 61 LRDLAAVEVPYFSDMEVKIITWGADTAACGDIISGLPVSAARRREIILGPADNFEQGW 120
QY 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 195
DB 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVFH 180
QY 196 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 255
DB 181 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVIPIVR 240
QY 256 RRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315
DB 241 RRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300
QY 316 TMR 318
DB 301 TMR 303

```

Search completed: May 6, 2004, 09:43:18

Job time: 36.167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 41.2434 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-10

Sequence: 1 AGITKVPYFPAQGLIRACM.....RGYAKAVDFPVSMTTMR 303

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1589	100.0	303	ABG32183	ABG32183 HCV prote
2	1589	100.0	334	ABG32182	ABG32182 HCV prote
3	1589	100.0	341	ABG32187	ABG32187 HCV prote
4	1589	100.0	352	ABG32186	ABG32186 HCV prote
5	1589	100.0	380	ABG32185	ABG32185 HCV prote
6	1589	100.0	393	ABG32184	ABG32184 HCV prote
7	1589	100.0	409	ABG32181	ABG32181 HCV prote
8	1580	99.4	303	ABG32191	ABG32191 HCV prote
9	1579	99.4	303	ABG32189	ABG32189 HCV prote
10	1570	98.8	301	ABG32190	ABG32190 HCV prote
11	1532	96.4	292	ABG32188	ABG32188 HCV prote
12	1530	96.3	2201	ABG30601	ABG30601 Hepatitis
13	1530	96.3	2201	ABG30591	ABG30591 Hepatitis
14	1530	96.3	2201	ABG30600	ABG30600 Hepatitis
15	1530	96.3	2201	ABG30581	ABG30581 Hepatitis
16	1530	96.3	2201	ABG30593	ABG30593 Hepatitis
17	1530	96.3	2201	ABG30582	ABG30582 Hepatitis
18	1530	96.3	2201	ABG30580	ABG30580 Hepatitis
19	1530	96.3	2201	ABG30587	ABG30587 Hepatitis
20	1530	96.3	2201	ABG30599	ABG30599 Hepatitis
21	1530	96.3	2201	ABG30594	ABG30594 Hepatitis
22	1530	96.3	2201	ABG30598	ABG30598 Hepatitis
23	1530	96.3	2201	ABG30595	ABG30595 Hepatitis
24	1530	96.3	3010	ABG32458	ABG32458 Hepatitis
25	1530	96.3	3010	ABG32459	ABG32459 Hepatitis

26	1530	96.3	3010	5	ABG32451	Abg32451 Hepatitis
27	1530	96.3	3010	5	ABG32455	Abg32455 Hepatitis
28	1530	96.3	3010	5	ABG32457	Abg32457 Hepatitis
29	1530	96.3	3010	5	ABG32460	Abg32460 Hepatitis
30	1530	96.3	3010	5	ABG32461	Abg32461 Hepatitis
31	1530	96.3	3010	5	ABG32454	Abg32454 Hepatitis
32	1530	96.3	3010	5	AAE20477	AAE20477 HCV-S1 fu
33	1530	96.3	3011	5	ABG32456	Abg32456 Hepatitis
34	1527	96.1	2201	5	ABG30586	Abg30586 Hepatitis
35	1527	96.1	2201	5	ABG30589	Abg30589 Hepatitis
36	1527	96.1	2201	5	ABG30583	Abg30583 Hepatitis
37	1527	96.1	2201	5	ABG30588	Abg30588 Hepatitis
38	1527	96.1	2307	3	AA770064	AA770064 Recombina
39	1527	96.1	3010	2	AAE86622	AAE86622 HCV prote
40	1527	96.1	3010	2	AAE82694	AAE82694 Partial H
41	1526	96.0	2201	5	ABG30590	Abg30590 Hepatitis
42	1524	95.9	2307	3	AA770065	AA770065 Recombina
43	1524	95.9	3010	5	ABG32452	Abg32452 Hepatitis
44	1523	95.8	2201	5	ABG30584	Abg30584 Hepatitis
45	1523	95.8	2201	5	ABG30602	Abg30602 Hepatitis

ALIGNMENTS

RESULT 1  
ABG32183  
ID ABG32183 standard; protein; 303 AA.

AC ABG32183 ;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 904-1206.  
XX  
XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
KW chronic liver disease; cirrhosis; end-stage liver disease; virologic;  
KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KW chaotropic agent; mutant; muten.  
XX  
XX Hepatitis C virus.  
OS Synthetic.  
OS  
PN WO200248375-A2.  
XX  
PD 20-JUN-2002.  
XX  
PF 13-DEC-2001; 2001WO-CA001796.  
XX  
PR 15-DEC-2000; 2000US-0256031P.  
XX

PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
XX  
XX Claim 39; Page 58-59; 67pp; English.  
PS

XX The invention relates to an isolated polypeptide consisting of a full-  
length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32182; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded, inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 904-1206  
 CC (numbered relative to the full length NS2/3 protein)

SQ Sequence 303 AA;

Query Match 100.0%; Score 1589; DB 5; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVPAQGLIRACMLVKKRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60  
 DB 1 AGITKVPYFVPAQGLIRACMLVKKRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60  
 QY 61 LRLDAVAVEPIFSDMEVKIITWGAADTPAACGDIISGLPVASARRGREILLGPADNFEQGM 120  
 DB 61 LRLDAVAVEPIFSDMEVKIITWGAADTPAACGDIISGLPVASARRGREILLGPADNFEQGM 120  
 QY 121 RLAPITAYSQQTRFGLIGCIITSLTGRDKNOVGCVQVNSTANOSPLATCVNVCMTVH 180  
 DB 121 RLAPITAYSQQTRFGLIGCIITSLTGRDKNOVGCVQVNSTANOSPLATCVNVCMTVH 180  
 QY 181 GAGSKTAGPRGPITQMTVNDODLVGMOAPFGASMTPTCGSSDLYLTRADYIPVR 240  
 DB 181 GAGSKTAGPRGPITQMTVNDODLVGMOAPFGASMTPTCGSSDLYLTRADYIPVR 240  
 QY 241 RRGDSRGSLLSPRVSYLKSGSGPILCPGSHAVGTRRAVCTRGVAKANDPFPVSMET 300  
 DB 241 RRGDSRGSLLSPRVSYLKSGSGPILCPGSHAVGTRRAVCTRGVAKANDPFPVSMET 300  
 QY 301 TMR 303  
 DB 301 TMR 303

RESULT 2

ABG32182 ID ABG32182 standard; protein; 334 AA.

AC ABG32182;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.

KM HCV, enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;

KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;

KM hepatotropic; anti-inflammatory; lauryldiethylamine oxide; LDAO;

KM chaotropic agent; 4K-6H (904-1206)st-4K; mutant; muten.

XX Hepatitis C virus.

OS Synthetic.

EH Key Location/Qualifiers  
 FT Peptide 1..15  
 FT /note= "4-Lys/His tag"  
 FT Protein 16..302  
 FT /note= "Truncated NS2/3 protease"  
 FT Peptide 319..334  
 FT /note= "Streptavidin/4-Lys tag"

W0200248375-A2.

20-JUN-2002.

13-DEC-2001; 2001WO-CA001796.

15-DEC-2000; 2000US-0256031P.

(BOHR) BOEHRINGER INGELHEIM CANADA LTD.

Thibault D, Lamare D, Maurice R, Pilote L, Pause A;

WPI, 2002-599511/64.

N-PSDB; ABR90407.

Novel polypeptide for screening inhibitors of non-structural proteases  
 useful as therapeutic agents against hepatitis C virus, comprises full  
 length non-structural protease, or its truncation.

Claim 39; Fig 9B; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32182; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the isolated protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 4K-6H (904-  
 CC 1206)st-4K comprising a truncated NS2/3 protein with a four Lys/six His N  
 CC-terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag

SQ Sequence 334 AA;

Query Match 100.0%; Score 1589; DB 5; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVPAQGLIRACMLVKKRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 60  
 DB 1 AGITKVPYFVPAQGLIRACMLVKKRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAG 75

QY 61 LRLDLAVAVEPVIFSPMEVKIITWGADTAACGDIISGLPVASARRGRILLGPADNFEQGM 120  
 Db 76 LRLDLAVAVEPVIFSPMEVKIITWGADTAACGDIISGLPVASARRGRILLGPADNFEQGM 135  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVPH 180  
 Db 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVPH 195  
 QY 181 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 240  
 Db 196 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 255  
 QY 241 RRGDSRGSILSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 256 RRGDSRGSILSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 QY 301 TMR 303  
 Db 316 TMR 318

## RESULT 3

ABG32187  
 ID ABG32187 standard; protein; 341 AA.

AC ABG32187;  
 XX

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 866-1206.

XX HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;

KM chronic liver disease; cirrhosis; end-stage liver disease; virucide;

KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;

KW chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

PN WO200248375-A2.

XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.

XX 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases

XX useful as therapeutic agents against hepatitis C virus, comprises full

XX length non-structural protease, or its truncation.

XX Claim 41; Page 62-63; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a

CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 866-1206  
 CC (numbered relative to the full length NS2/3 protein)

SO Sequence 341 AA;

Query Match 100.0%; Score 1589; DB 5; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 66-146;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVHTPLQDMAHAG 60  
 Db 39 AGITKVPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVHTPLQDMAHAG 98  
 QY 61 LRLDLAVAVEPVIFSPMEVKIITWGADTAACGDIISGLPVASARRGRILLGPADNFEQGM 120  
 Db 99 LRLDLAVAVEPVIFSPMEVKIITWGADTAACGDIISGLPVASARRGRILLGPADNFEQGM 158  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVPH 180  
 Db 159 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVPH 218  
 QY 181 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 240  
 Db 219 GAGSKTLAGKPGPTQMTVNDQLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 278  
 QY 241 RRGDSRGSILSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 279 RRGDSRGSILSPRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 338  
 QY 301 TMR 303  
 Db 339 TMR 341

RESULT 4  
 ABG32186  
 ID ABG32186 standard; protein; 352 AA.  
 XX  
 AC ABG32186;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 truncation mutant 855-1206.  
 XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
 KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KW chaotropic agent; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 PN WO200248375-A2.  
 XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 61-62; 67pp; English.  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 855-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 CC XX  
 CC Sequence 352 AA;  
 CC  
 CC Query Match 100.0%; Score 1589; DB 5; Length 352;  
 CC Best Local Similarity 100.0%; Pred. No. 6, 3e-146;  
 CC Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 1 AGTGTGVTYRAGGLTRACMLVRKAGGHHYQMAFMKALALTTYYDDHTPQDMANAG 60  
 CC 50 AGITKQVYFVRAQGLTRACMLVRKAGGHHYQMAFMKALALTTYYDDHTPQDMANAG 109  
 CC 61 LRLDAVAVBEVITSDMEVKITITGADTRACGDIISGLPVASRGREILIGPANFEGQGM 120  
 CC 110 LRLDAVAVBEVITSDMEVKITITGADTRACGDIISGLPVASRGREILIGPANFEGQGM 169  
 CC 121 RLAPITAYSQQTRGLIGCIITSLTGRDKNOVEGEVQVNSTATQSFPLATCVNVCVTVFH 180  
 CC 170 RLAPITAYSQQTRGLIGCIITSLTGRDKNOVEGEVQVNSTATQSFPLATCVNVCVTVFH 229  
 CC 181 GAGSKTLAAGKRPITQWNTWVDIVGMOAPPPARASMTPTCGSSDLVYTRADVIYPR 240  
 CC 230 GAGSKTLAAGKRPITQWNTWVDIVGMOAPPPARASMTPTCGSSDLVYTRADVIYPR 289

QY 241 RRGDSRGSLSPRVSYLKGSGGFLPCSGHAGVIFRAAVCTRGVAKAVDPIPVESMET 300  
 DB 290 RRGDSRGSLSPRVSYLKGSGGFLPCSGHAGVIFRAAVCTRGVAKAVDPIPVESMET 349  
 QY 301 TMR 303  
 DB 350 TMR 352  
 RESULT 5  
 ABG32185  
 ID ABG32185 standard; protein, 360 AA.  
 AC ABG32185;  
 XX 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation mutant 827-1206.  
 DB HCV, enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; viruslike;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutcin.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 PN WO200248375-A2.  
 PD 20-JUN-2002.  
 PF 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 60-61; 67pp; English.  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is

CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 CC  
 XX Sequence 380 AA;

Query Match 100.0%; Score 1589; DB 5; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 7e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALGTGYVDHLTPLODMANAG 60  
 Db 78 AGITKVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALGTGYVDHLTPLODMANAG 137  
 QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARRGREILLGPADNFEQGW 120  
 Db 138 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARRGREILLGPADNFEQGW 197  
 QY 121 RLAPITAVSQOTRGLICITISLTGRDKQVGEVGVVSTATOSFLATCVNGVCTVFFH 180  
 Db 198 RLAPITAVSQOTRGLICITISLTGRDKQVGEVGVVSTATOSFLATCVNGVCTVFFH 257  
 QY 181 GAGSKTLGAPKGPITQMTYTNVDQLVGMQAPPGARSMTPTCGSSDLVLRHADVIPIVR 240  
 Db 258 GAGSKTLGAPKGPITQMTYTNVDQLVGMQAPPGARSMTPTCGSSDLVLRHADVIPIVR 317  
 QY 241 RRGDSRGSLSRPVSYLKSGSGGPLICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 318 RRGDSRGSLSRPVSYLKSGSGGPLICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 377  
 QY 301 TMR 303  
 Db 378 TMR 380

RESULT 6  
 ABG32184  
 ID ABG32184 standard; protein; 393 AA.  
 XX  
 AC ABG32184;

XX 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 truncation mutant 815-1206.

XX HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 XX hepatocellular carcinoma; anti-inflammatory; lauryldiethyamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutagen.

XX Hepatitis C virus.  
 OS Synthetic.

XX W0200248375-A2.

XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.

XX 15-DEC-2000; 2000US-0256031P.

XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-559511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full

PT length non-structural protease, or its truncation.  
 XX  
 XX Claim 41; Page 59-60; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethyamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 CC  
 XX Sequence 393 AA;

Query Match 100.0%; Score 1589; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITVVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALGTGYVDHLTPLODMANAG 60  
 Db 91 AGITVVPYFVRAOGLIRACMLVRKAGGHVQMAFMKLAALGTGYVDHLTPLODMANAG 150  
 QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARRGREILLGPADNFEQGW 120  
 Db 151 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARRGREILLGPADNFEQGW 210  
 QY 121 RLAPITAVSQOTRGLICITISLTGRDKQVGEVGVVSTATOSFLATCVNGVCTVFFH 180  
 Db 211 RLAPITAVSQOTRGLICITISLTGRDKQVGEVGVVSTATOSFLATCVNGVCTVFFH 270  
 QY 181 GAGSKTLGAPKGPITQMTYTNVDQLVGMQAPPGARSMTPTCGSSDLVLRHADVIPIVR 240  
 Db 271 GAGSKTLGAPKGPITQMTYTNVDQLVGMQAPPGARSMTPTCGSSDLVLRHADVIPIVR 330  
 QY 241 RRGDSRGSLSRPVSYLKSGSGGPLICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 331 RRGDSRGSLSRPVSYLKSGSGGPLICPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 390  
 QY 301 TMR 303  
 Db 391 TMR 393

RESULT 7  
 ABG32181  
 ID ABG32181 standard; protein; 409 AA.  
 XX

AC ABG32181;  
 XX 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 (810-1206).  
 DE HCV protease NS2/3 (810-1206).  
 XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 XX hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT Peptide 398..409  
 FT /note="Streptavidin tag"  
 XX WO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 DR N-PSDB; ABK90406.  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX Claim 42; Fig 1B; 67pp; English.  
 PS The invention relates to an isolated polypeptide consisting of a full-  
 XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) ID-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 (810-1206) protein, which has a C-  
 CC terminal streptavidin tag

SQ Sequence 409 AA;  
 Query Match 100.0%; Score 1589; DB 5; Length 409;  
 Best local similarity 100.0%; Pred. No. 7, 8e-146;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGITKVFYFPAQGLIRACMLVRKAGHYVQVAFMKLALTGTYVDHLLTPLOMAHAG 60  
 DB 95 AGITKVFYFPAQGLIRACMLVRKAGHYVQVAFMKLALTGTYVDHLLTPLOMAHAG 154  
 QY 61 LRDLAVAVEPITPSDMVKIITWQADTAACGDIISGLPVARRGREILLGPANFEGQW 120  
 DB 155 LRDLAVAVEPITPSDMVKIITWQADTAACGDIISGLPVARRGREILLGPANFEGQW 214  
 QY 121 RLAPITAVSQOTGLGCIITSLTGRDKNQVEGVQVSTATQSPFLATCVNGVCTVFF 180  
 DB 215 RLAPITAVSQOTGLGCIITSLTGRDKNQVEGVQVSTATQSPFLATCVNGVCTVFF 274  
 QY 181 GAGSKTLAGPKPITQWYTNVDQVLVGMQAPPGARSMTPCTCGSSDLIYLTRADVI 240  
 DB 275 GAGSKTLAGPKPITQWYTNVDQVLVGMQAPPGARSMTPCTCGSSDLIYLTRADVI 334  
 QY 241 RRGDSRGSILSPRVSYLKSSGGPILCPGSHAVGIPRAVCTRGVAKAVDFIPVESMET 300  
 DB 335 RRGDSRGSILSPRVSYLKSSGGPILCPGSHAVGIPRAVCTRGVAKAVDFIPVESMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397  
 RESULT 8  
 ABG32191  
 ID ABG32191 standard; protein; 303 AA.  
 AC ABG32191;  
 XX 05-NOV-2002 (first entry)  
 DT HCV protease NS2/3 truncation 904-1206/Cys9393Aa.  
 DE HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 XX hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT Misc-difference 90 /note="Wild-type Cys substituted by Ala"  
 XX WO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 DR Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Disclosure; Page 65-66; 67pp; English.



CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation of a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent, and a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterization of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant Cys993Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

XX Sequence 303 AA;

Query Match 99.4%; Score 1580; DB 5; Length 303;  
Best Local Similarity 99.7%; Pred. No. 3.8e-145;  
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGITKVPYFRAQGLIAPACMLVKAAGHYVQAEMKLAALTGYVDHITPLQDMAHAG 60  
DB 1 AGITKVPYFRAQGLIAPACMLVKAAGHYVQAEMKLAALTGYVDHITPLQDMAHAG 60  
QY 61 LRLDAVAVEPVIFSDMEVKIITWGADTFAAGSDIISGLPVASRRGRRELLGPANPFEGQGM 120  
DB 61 LRLDAVAVEPVIFSDMEVKIITWGADTFAAGSDIISGLPVASRRGRRELLGPANPFEGQGM 120  
QY 121 RLAPITAYVQQQTRGLIGCIITSLTGSDKNQVGEVAVSTATQSFPLATCVNGCVCTVPH 180  
DB 121 RLAPITAYVQQQTRGLIGCIITSLTGSDKNQVGEVAVSTATQSFPLATCVNGCVCTVPH 180  
QY 181 GAGSKTLAGKSPITQWYTNVDDIVGMQAPPGARSPSTPCTCGSSDLYLTRADVIIPR 240  
DB 181 GAGSKTLAGKSPITQWYTNVDDIVGMQAPPGARSPSTPCTCGSSDLYLTRADVIIPR 240  
QY 241 RRGDSRGSLLSPRPVSTLKSGSSGPLLCPGHAHVGFRAVACRGYAKAVDFIPVSMKT 300  
DB 241 RRGDSRGSLLSPRPVSTLKSGSSGPLLCPGHAHVGFRAVACRGYAKAVDFIPVSMKT 300  
QY 301 TMR 303  
DB 301 TMR 303

RESULT 9  
ABG32189  
ID ABG32189 standard; protein; 303 AA.  
XX ABG32189;  
XX 05-NOV-2002 (first entry)  
DT

XX HCV protease NS2/3 truncation 904-1206/His952Ala.  
DE HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; muten.  
XX Hepatitis C virus.  
OS Synthetic.  
OS Key Location/Qualifiers  
FH Misc-difference 45 /note= "wild-type His substituted by Ala"  
FT MO200248375-A2.  
XX 20-JUN-2002.  
XX 13-DEC-2001; 2001WO-CA001796.  
XX 15-DEC-2000; 2000US-0256031P.  
XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
XX Example 7, Fig 8, 67pp; English.

PS The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent, and a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterization of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant His952Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

Query Match 99.4%; Score 1579; DB 5; Length 303;  
XX Sequence 303 AA;

Best Local Similarity 99.7%; Pred. No. 4.8e-145;  
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGITKPYFVFAAGLIRACMLVRKAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAG 60  
DB 1 AGITKPYFVFAAGLIRACMLVRKAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAG 60  
QY 61 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRREILLGPADNFEQGM 120  
DB 61 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRREILLGPADNFEQGM 120  
QY 121 RLIAPIITAVSQOTRGLGCIITSLTGDRDNQVEGEVQVSTATQSPFLATCVNGVMTVFH 180  
DB 121 RLIAPIITAVSQOTRGLGCIITSLTGDRDNQVEGEVQVSTATQSPFLATCVNGVMTVFH 180  
QY 181 GAGSKTLAGPKGPITQMTNTVDODLVGWOAPPGARSMTPTCGSSDLVLTSHADVIPIVR 240  
DB 181 GAGSKTLAGPKGPITQMTNTVDODLVGWOAPPGARSMTPTCGSSDLVLTSHADVIPIVR 240  
QY 241 RRGDSRGSLLSPRPVSYLKSGSGGFLCPSGHAIVGIFRAAVCTRGVAKAVDFIIVESMET 300  
DB 241 RRGDSRGSLLSPRPVSYLKSGSGGFLCPSGHAIVGIFRAAVCTRGVAKAVDFIIVESMET 300  
QY 301 TMR 303  
DB 301 TMR 303

RESULT 10  
ABG32190  
ID ABG32190 standard; protein; 301 AA.

AC ABG32190;  
DT 05-NOV-2002 (first entry)  
DE HCV protease NS2/3 truncation 904-1206/deltaLeu1026-1A1027.

XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; virologic;  
XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
XX chaotropic agent; mutant; mutuin.

OS Hepatitis C virus.  
XX Synthetic.

XX Key Location/Qualifiers  
FT Misc-difference 122..123  
FT /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"

XX WO200248375-A2.

XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CAD001796.

XX 15-DEC-2000; 2000US-0256031P.

XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thiobault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.

XX Example 7; Page 64-65; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal

CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1p-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, and LDAO in the presence of reduced concentration of the  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation 904-1206 mutant  
CC deltaLeu1026-1A1027 (numbered relative to the full length NS2/3 protein)  
CC a mutant devoid of autocatalytic activity

XX Sequence 301 AA;

XX Query Match 98.8%; Score 1570; DB 5; Length 301;

XX Best Local Similarity 99.3%; Pred. No. 3.6e-144;  
XX Matches 301; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 AGITKPYFVFAAGLIRACMLVRKAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAG 60  
DB 1 AGITKPYFVFAAGLIRACMLVRKAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAG 60  
QY 61 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRREILLGPADNFEQGM 120  
DB 61 LRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRREILLGPADNFEQGM 120  
QY 121 RLIAPIITAVSQOTRGLGCIITSLTGDRDNQVEGEVQVSTATQSPFLATCVNGVMTVFH 180  
DB 121 RLIAPIITAVSQOTRGLGCIITSLTGDRDNQVEGEVQVSTATQSPFLATCVNGVMTVFH 180  
QY 181 GAGSKTLAGPKGPITQMTNTVDODLVGWOAPPGARSMTPTCGSSDLVLTSHADVIPIVR 240  
DB 181 GAGSKTLAGPKGPITQMTNTVDODLVGWOAPPGARSMTPTCGSSDLVLTSHADVIPIVR 240  
QY 241 RRGDSRGSLLSPRPVSYLKSGSGGFLCPSGHAIVGIFRAAVCTRGVAKAVDFIIVESMET 300  
DB 241 RRGDSRGSLLSPRPVSYLKSGSGGFLCPSGHAIVGIFRAAVCTRGVAKAVDFIIVESMET 300  
QY 301 TMR 303  
DB 301 TMR 303

XX RESULT 11

XX ABG32188  
XX ID ABG32188 standard; protein; 292 AA.

XX AC ABG32188;

XX DT 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 915-1206.

KW HCV; enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection;  
 KV chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
 KW hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDXO;  
 KM chaotropic agent; mutant; mutein.  
 XX  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX  
 XX WO200248375-A2.  
 PN  
 XX  
 XX 20-JUN-2002.  
 PD  
 XX  
 XX 13-DEC-2001; 2001WO-CAN001796.  
 PF  
 XX  
 XX 15-DEC-2000; 2000US-0256031P.  
 PR  
 XX  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA  
 XX  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 PI  
 XX  
 XX WPI; 2002-599511/64.  
 XR  
 XX  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 XX useful as therapeutic agents against hepatitis C virus, comprises full  
 XX length non-structural protease, or its truncation.  
 XX  
 XX Claim 41, Page 63, 67pp, English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-12061)) or its truncation, having as its N-terminal residue amino acid 810 to 966, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldimethylamine oxide (DDAO) to prevent auto-cleavage of the protease, (2) a NS2/3 inhibitory peptide appearing as HGS32198, (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and DDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 915-1206 (numbered relative to the full length NS2/3 protein)

Sequence 292 AA;

Query Match	96.4%;	Score 1532;	DB 5;	Length 292;
Best Local Similarity	100.0%;	Pred. No. 1.7e-140;		
Matches 292;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 12 AGGLRACLTARKAAGHHVQMAFKKALGTYVYTHLPDQMAHAGRDIAVAVEPV 71

Db 1 AGGLRACLTARKAAGHHVQMAFKKALGTYVYTHLPDQMAHAGRDIAVAVEPV 60

QY 72 IFSDEVKLIITWGADTAACDPIISGLPYASABRGREILLGPADNPEQGKRLAPITAYSQ 131

Db 61 |FSMEVKITTWGADPAACGDIISGHPVARSGRFELLGPDNFEGQCMRLAVITAYSQ 120

QY 132 |QTEGLGCIITSLTGRDNQVGEVQVNSTAQSLATCVCVCTVTHGAGSKTLGPK 191

Db 121 |QTEGLGCIITSLTGRDNQVGEVQVNSTAQSLATCVCVCTVTHGAGSKTLGPK 180

QY 192 |GPTQWTVNDQVLVMOAPPGARSMTPCTGSSDLVTVTHSHADIVPRRAGDSRGLLS 251

Db 181 |GPTQWTVNDQVLVMOAPPGARSMTPCTGSSDLVTVTHSHADIVPRRAGDSRGLLS 240

QY 252 |PRPVSYLKSGSGPILLCPSGHAVGIFRAAVCTRGVAKAVDPIPVESMETTR 303

Db 241 |PRPVSYLKSGSGPILLCPSGHAVGIFRAAVCTRGVAKAVDPIPVESMETTR 292

XX	RESULT 12
PF	ABG30601
ID	ABG30601 standard; protein; 2201 AA.
AC	ABG30601;
XX	
DT	21-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS2/3, NS3/4, NS3 and NSSB mutant #10.
KM	Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX	cell culture replication; NS2/3; NS3/4; NS3; NSSB; mutant; mutein.
OS	Hepatitis C virus.
XX	Synthetic.
FH	Key Location/Qualifiers
FT	Misc-difference 882
FT	/label= Arg, Lys
FT	Misc-difference 2183
FT	/note= "Wild type Met substituted by Thr"
PN	WO200252015-A2.
XX	
XX	04-JUL-2002.
PD	
XX	
XX	20-DEC-2001; 2001MO-CA001843.
PF	
XX	
PR	22-DEC-2000; 2000US-0257857P.
XX	
PA	(BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.
XX	
PI	Kukoly G, Pause A;
PT	WPI; 2002-575382/61.
XX	
XX	
PS	
XX	
XX	
CC	Claim 3; Page; 140pp; English.
CC	The invention describes a self-replicating hepatitis C virus (HCV)
CC	polynucleotide molecule comprising a 5'-non translated region (NTR),
CC	where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC	region coding for a HCV polyprotein, and a 3'-NTR region. The self-
CC	replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC	potential inhibitors of HCV replication. The HCV RNA molecule is also
CC	useful for efficiently establishing cell culture replication. The self-
CC	replicating polynucleotide molecule contains a 5'-NTR, where G at
CC	position 1 is substituted for A, and therefore provides an alternative to
CC	existing systems comprising a self-replicating HCV RNA molecule that, in
CC	conjunction with mutations in the HCV non-structural region, such as the
CC	G(2042)/C/R mutations, transduces and/or replicates with greater
CC	efficiency. This amino acid sequence represents a mutant of the hepatitis
CC	C virus replicon Apkx12 and contains the viral protease NS2/3, protease
CC	complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NSSB. Note:

CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

CC Sequence 2201 AA;

Query Match 96.3%; Score 1530; DB 5; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFAHQGLIRACMLVRKAGHYVQNAFMKLAALTGYVVDHITPLQDWAHAG 60  
 DB 95 AGITKVPYFAHQGLIRACMLVRKAGHYVQNAFMKLAALTGYVVDHITPLQDWAHAG 154  
 QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 120  
 DB 155 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 214  
 QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVH 180  
 DB 215 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVH 274  
 QY 181 GAGSKTLAAGPKGPIITQMTYTNVDOLVGMQAPPGARSMTCTCGSSDLVLTTRHADVIPVR 240  
 DB 275 GAGSKTLAAGPKGPIITQMTYTNVDOLVGMQAPPGARSMTCTCGSSDLVLTTRHADVIPVR 334  
 QY 241 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESMET 300  
 DB 335 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397

RESULT 13

ABG30591 ID ABG30591 standard; protein; 2201 AA.

AC ABG30591;

DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for

PT evaluating potential inhibitors of HCV replication.

XX Claim 3; Page: 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apgk2 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

CC Sequence 2201 AA;

Query Match 96.3%; Score 1530; DB 5; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFAHQGLIRACMLVRKAGHYVQNAFMKLAALTGYVVDHITPLQDWAHAG 60  
 DB 95 AGITKVPYFAHQGLIRACMLVRKAGHYVQNAFMKLAALTGYVVDHITPLQDWAHAG 154  
 QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 120  
 DB 155 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARSGREIILGPADNPEGQW 214  
 QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVH 180  
 DB 215 RLAPITAVSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCWTVH 274  
 QY 181 GAGSKTLAAGPKGPIITQMTYTNVDOLVGMQAPPGARSMTCTCGSSDLVLTTRHADVIPVR 240  
 DB 275 GAGSKTLAAGPKGPIITQMTYTNVDOLVGMQAPPGARSMTCTCGSSDLVLTTRHADVIPVR 334  
 QY 241 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESMET 300  
 DB 335 RRGDSRGSLLSPRPVSYLKSGSGGPIICPSGHAAGIFRAAVCTRGVAKAVDFPVESMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397

RESULT 14

ABG30600 ID ABG30600 standard; protein; 2201 AA.

AC ABG30600;

DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

OS Synthetic.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for

Misc-difference 1357  
/note= "Wild type Pro substituted by Leu"  
W0200252015-A2.  
04-JUL-2002.  
20-DEC-2001; 2001WO-CA001843.  
22-DEC-2000; 2000US-0257857P.  
(BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.  
Kukolj G, Pause A;  
WPI; 2002-575382/61.  
New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.

Claim 3; Page: 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polypeptide region coding for a HCV polypeptide; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon Apgk12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention

Sequence 2201 AA;

Query Match 96.3%; Score 1530; DB 5; Length 2201;  
Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

1 AGITKVPYFVRAAGLIRACMLVRKAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAG 60  
95 AGITKVPYFVRAAGLIRACMLVRKAAGHYVQMALMLALGTGYVDHLTPLODMAHAG 154  
61 LRDLAFAVEPVIFSDMEVKITWGADTAACGDIISGLPVSAARREIILGPADNFGQGM 120  
155 LRDLAFAVEPVIFSDMEVKITWGADTAACGDIISGLPVSAARREIILGPADNFGQGM 214  
121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVPH 180  
215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVPH 274  
181 GAGSKTLAAGPKGPIITOMYTNVDOLVGWQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 240  
275 GAGSKTLAAGPKGPIITOMYTNVDOLVGWQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 334  
241 RRGSGRGLSLPRVSYLKSSGGPLCPGSHAVGIFRAAVCTGVAKADVFVVESEMET 300  
335 RRGSGRGLSLPRVSYLKSSGGPLCPGSHAVGIFRAAVCTGVAKADVFVVESEMET 394  
301 TMR 303  
395 TMR 397

RESULT 15

ABG30581 standard; protein; 2201 AA.

AC ABG30581;

DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.

KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NS5B.

OS Hepatitis C virus.

PN W0200252015-A2.

PD 04-JUL-2002.

PF 20-DEC-2001; 2001WO-CA001843.

PR 22-DEC-2000; 2000US-0257857P.

PA (BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.

Kukolj G, Pause A;

WPI; 2002-575382/61.

N-PSDB; ABK88573.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.

PS Disclosure; Page 49-58; 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polypeptide region coding for a HCV polypeptide; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations, transduces and/or replicates with greater efficiency. This amino acid sequence is encoded by the hepatitis C virus replicon Apgk12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

Sequence 2201 AA;

Query Match 96.3%; Score 1530; DB 5; Length 2201;  
Best Local Similarity 95.0%; Pred. No. 4.6e-139;  
Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

1 AGITKVPYFVRAAGLIRACMLVRKAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAG 60  
95 AGITKVPYFVRAAGLIRACMLVRKAAGHYVQMALMLALGTGYVDHLTPLODMAHAG 154  
61 LRDLAFAVEPVIFSDMEVKITWGADTAACGDIISGLPVSAARREIILGPADNFGQGM 120  
155 LRDLAFAVEPVIFSDMEVKITWGADTAACGDIISGLPVSAARREIILGPADNFGQGM 214  
121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVPH 180  
215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVPH 274  
181 GAGSKTLAAGPKGPIITOMYTNVDOLVGWQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 240  
275 GAGSKTLAAGPKGPIITOMYTNVDOLVGWQAPPGARSMTPTCTGSSDLYVTRHADVIPVR 334

Qy	241	RRGDSRGSLSPRPVSTLKSSGGPLICPSGHA	VGIFRAAVCTRGVAKAVDFIVESMET	300
Db	335	RRGDSRGSLSPRPVSTLKSSGGPLICPSGHA	VGIFRAAVCTRGVAKAVDFIVESMET	394
Qy	301	TMR	303	
Db	395	TMR	397	

Search completed: May 6, 2004, 09:30:44  
Job time : 42.243 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 9.86806 Seconds

(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-10

Perfect score: 1589

Sequence: 1 AGITKVPYFVRAQGLIRACM.....RGVAKAVDFIPVESMETTNR 303

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.78.\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1532	96.4	3010	1 A45573	genome polyprotein
2	1527	96.1	3010	1 GNMVVC	genome polyprotein
3	1514	95.3	3010	1 GNMVTC	genome polyprotein
4	1486	93.5	3010	1 S18030	genome polyprotein
5	1478	93.0	3010	1 GNMVTC	genome polyprotein
6	1403	88.3	3011	1 S40770	genome polyprotein
7	1398	88.0	3011	1 GNMVVC	genome polyprotein
8	1385	87.2	3011	1 GNMVTC	genome polyprotein
9	1234	77.7	3014	1 UC5620	genome polyprotein
10	1172	73.8	3033	1 JQ1303	genome polyprotein
11	1150	72.4	3033	1 GNMVTC	genome polyprotein
12	397.5	25.0	3005	2 T08841	genome polyprotein
13	341	21.5	2970	2 T08839	polyprotein - mam
14	101	6.4	600	2 B46642	DNA-directed DNA p
15	99.5	6.3	353	2 G87382	conserved hypochet
16	97.5	6.1	1085	2 T03531	cohn protein homol
17	95.5	6.0	470	2 UC4098	tetracycline 6-hyd
18	93	5.9	660	2 VHMW12	structural protein
19	92.5	5.8	706	2 S33761	transferrin precu
20	92.5	5.8	716	2 G83612	hypothetical prote
21	91	5.7	904	2 A84212	hypothetical prote
22	90.5	5.7	868	2 H81775	aconitate hydratase
23	90	5.7	2796	2 UC4743	fatty-acid synthase
24	89.5	5.6	659	2 T36248	CDA peptidic synthe
25	88	5.5	3414	1 B44212	structural protein
26	87	5.4	3412	1 GNMVTC	genome polyprotein
27	86.5	5.4	3412	1 NMIVW8	genome polyprotein
28	85.5	5.4	348	2 H70549	exo-alpha-sialidase
29					probable pnhb proc

#### ALIGNMENTS

##### RESULT 1

A45573 genome polyprotein - hepatitis C virus (strain UT)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

C:Accession: A45573

R/Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, N.

Virus Res. 23, 39-53, 1992

A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: se

A:Reference number: A45573; PMID:92295714; PMID:1318627

A:Accession: A45573

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-3010 <TAN>

A/Cross-references: GB:ID1168; GB:DOI1171; NID:G221612; PID:BA01943.1; PID:G221613

A/Experimental source: HCV-UT

A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polyprotein; serine

F:115-191/Product: capsid protein C #status predicted <CPC>

F:192-389/Product: envelope protein M #status predicted <EM>

F:390-729/Product: major envelope protein B #status predicted <MB>

F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>

F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NA>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NB>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS>

Query Match 96.4% Score 1532; DB 1; Length 3010;  
Best Local Similarity 95.4% Pred. No. 7.3e-12;  
Matches 289; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY	1	AGITKVPYFVRAQGLIRACMVRKAGHYVQMAFMKLAALTGIVYDHLTFLQDMAHAG 60	
DB	904	AAITAMPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGIVYDHLTFLQDMAHAG 963	
QY	61	LELDLVAVEPVFESMEKRIITWGDYACGIIIGLVSARRGREIILGPADNIEGGGW 120	
DB	964	LELDLVAVEPVFESMEKRIITWGDYACGIIIGLVSARRGREIILGPADNIEGGGW 1023	
QY	121	RLAFLITAYSQGTRELLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMTFEH 180	
DB	1024	RLAFLITAYSQGTRELLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMTFEH 1093	
QY	181	GAGSKTLGPKPIQTQNTTNDQDLVWGQAPPGASMTPTCGSSDLYLVTRHADVIYVR 240	

Db 1084 GAGSKTLAAGPKPIITOMYTNVDDLVGMHAPPGARSLTFCTCGSSDLYLVTRHADVIPIVR 1143  
 |||  
 QY 241 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 |||  
 Db 1144 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||  
 QY 301 TMR 303  
 |||  
 Db 1204 TMR 1206

## RESULT 2

GNMVTM

genome polypeptide - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein E; major envelope protein B; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001

C:Accession: A39253; PS0086

R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimoto Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients

A:Reference number: A39253; MUID:1088550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 &lt;KAT&gt;

A:Cross-references: GB:D0208; NID:G221610; PID:BA44233.1; PID:G221611

R:Kato, N.; Okoshi, S.; Shimotohno, K. Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A:Reference number: PS0086

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 &lt;KAT&gt;

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polypeptide have not been determined.

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; setin

F/2-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F/116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F/192-389/Product: major envelope protein E #status predicted &lt;MEB&gt;

F/390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F/730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F/1007-1615/Product: hepatitis C virus predicted &lt;NS3&gt;

F/1230-1237/Product: nucleotide-binding motif A (P-loop)

F/1312-1317/Product: nucleotide-binding motif B

F/1316-1319/Product: DEHX motif

F/1616-1862/Product: nonstructural protein NS4a #status predicted &lt;NS4a&gt;

F/1863-2013/Product: nonstructural protein NS4b #status predicted &lt;NS4b&gt;

F/2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F/136,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 96.1%; Score 1527; DB 1; Length 3010;

Best Local Similarity 94.4%; Pred. No. 2e-123;

Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

Db 1144 RRDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 |||  
 QY 301 TMR 303  
 |||  
 Db 1204 TMR 1206

## RESULT 3

GNMVTM

genome polypeptide - hepatitis C virus (strain Taiwan)

N:Contains: capsid protein C; envelope protein E; hepatitis C virus (strain Taiwan) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001

C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S. Virology 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the

A:Reference number: A40244; MUID:9230206; PMID:1314449

A:Accession: A40244

A:Molecule type: genomic RNA

A:Residues: 1-3010 &lt;CHE&gt;

A:Cross-references: GB:M84754

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F/1-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F/116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F/192-389/Product: major envelope protein E #status predicted &lt;MEB&gt;

F/390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F/730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F/1007-1615/Product: hepatitis C virus predicted &lt;NS3&gt;

F/1230-1237/Product: nucleotide-binding motif A (P-loop)

F/1312-1317/Product: nucleotide-binding motif B

F/1316-1319/Product: DEHX motif

F/1616-1862/Product: nonstructural protein NS4a #status predicted &lt;NS4a&gt;

F/1863-2013/Product: nonstructural protein NS4b #status predicted &lt;NS4b&gt;

F/2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F/136,209,233,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,

Query Match 95.3%; Score 1514; DB 1; Length 3010;

Best Local Similarity 93.1%; Pred. No. 2.6e-122;

Matches 282; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 1 AGITVVFVFAOGLIRACMLVRKAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAG 60  
 |||  
 Db 904 AGITVVFVFAOGLIRACMLVRKAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAG 963  
 |||  
 QY 61 LRDLAFAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARRREIILGPADNFEQGW 120  
 |||  
 Db 964 LRDLAFAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARRREIILGPADNFEQGW 1023  
 |||  
 QY 121 RLAPITAYSOOTRGLICITITSLTRGDKVGEVGVVSTATOSFLATCINVCMTVPH 180  
 |||  
 Db 1024 RLAPITAYSOOTRGLICITITSLTRGDKVGEVGVVSTATOSFLATCINVCMTVPH 1083  
 |||  
 QY 181 GAGSKTLAAGPKPIITOMYTNVDDLVGMHAPPGARSLTFCTCGSSDLYLVTRHADVIPIVR 240  
 |||  
 Db 1084 GAGSKTLAAGPKPIITOMYTNVDDLVGMHAPPGARSLTFCTCGSSDLYLVTRHADVIPIVR 1143  
 |||  
 QY 301 TMR 303  
 |||  
 Db 1204 TMR 1206

## RESULT 4

GNMVTM

genome polypeptide - hepatitis C virus (isolate JKI)



N:Contains: capsid protein C; envelope protein M; hepatitis C virus (HCV 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C/Species: hepatitis C virus  
C/Variety: isolate JKI  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
C/Accession: S18030; S33570; A48332; S18029  
R/Honda, M.; Kaneko, S.; Masahashi, U.; Kobayashi, K.; Murakami, S.  
Submitted to the EMBL Data Library, September 1991  
A/Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient.  
A/Reference number: S18028  
A/Accession: S18030  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <HON>  
A/Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA3793.1; PID:G59479  
A/Experimental source: isolate JKI from an individual  
R/Honda, M.; Kaneko, S.; Uenura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
A/Reference number: A48332; MUID:33119270; PMID:8380322  
A/Accession: S33570  
A/Molecule type: genomic RNA  
A/Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
A/Cross-references: EMBL:X61591  
A/Note: this sequence is inconsistent with the nucleotide translation  
A/Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
A/Note: sequence extracted from NCBI database (NCBI:121747, NCBI:121748)  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitis C virus genome polyprotein  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif A (P-loop)  
F/1316-1319/Region: DEXH motif  
F/1616-1662/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS

Query Match 93.5%; Score 1486; DB 1; Length 3010;  
Best Local Similarity 92.4%; Pred. No. 76-120;  
Matches 280; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 1 AGITKVPFVRAQGLIRACMLVRKAGAGHYVQMAFMKLAALGTGYVDHLTPLOQMAHAG 60  
DB 904 AGITRVPFVRAQGLIRACMLVRKAGAGHYVQMAFMKLAALGTGYVDHLTPLOQMAHAG 963  
QY 61 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVSARGREILLGPADNFEQGM 120  
DB 964 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVSARGREILLGPADNFEQGM 1023  
QY 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 1083  
QY 181 GAASKTLAGKPGITQMTYTNVDQIVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPIVR 240  
DB 1084 GAASKTLAGKPGITQMTYTNVDQIVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPIVR 1143  
QY 241 RRDSRGLSPRPVSYLKSGSGPILCPGHAAGVIFPAVCTRGVAKAVDFIPVESMET 300  
DB 1144 RRDSRGLSPRPVSYLKSGSGPILCPGHAAGVIFPAVCTRGVAKAVDFIPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 5

GNMWTG  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (HCV 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C/Species: hepatitis C virus  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
C/Accession: A38465  
R/Takami, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
A/Title: Structure and organization of the hepatitis C virus genome isolated from human  
A/Reference number: A38465; MUID:91140698; PMID:1847440  
A/Accession: A38465  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <TKK>  
A/Cross-references: EMBL:M58335; NID:G329770; PIDN:AA72945.1; PID:G329771  
A/Experimental source: isolate JKI from an individual  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitis C virus genome polyprotein  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif A (P-loop)  
F/1316-1319/Region: DEXH motif  
F/1616-1662/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 93.0%; Score 1478; DB 1; Length 3010;  
Best Local Similarity 92.1%; Pred. No. 3-5e-113;  
Matches 279; Conservative 11; Mismatches 13; Indels 0; Gaps 0;

QY 1 AGITKVPFVRAQGLIRACMLVRKAGAGHYVQMAFMKLAALGTGYVDHLTPLOQMAHAG 60  
DB 904 AGITRVPFVRAQGLIRACMLVRKAGAGHYVQMAFMKLAALGTGYVDHLTPLOQMAHAG 963  
QY 61 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVSARGREILLGPADNFEQGM 120  
DB 964 LRDLAAVEVPFVSMEXKITTWGADTAACGDIISGLPVSARGREILLGPADNFEQGM 1023  
QY 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
DB 1024 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 1083  
QY 181 GAASKTLAGKPGITQMTYTNVDQIVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPIVR 240  
DB 1084 GAASKTLAGKPGITQMTYTNVDQIVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPIVR 1143  
QY 241 RRDSRGLSPRPVSYLKSGSGPILCPGHAAGVIFPAVCTRGVAKAVDFIPVESMET 300  
DB 1144 RRDSRGLSPRPVSYLKSGSGPILCPGHAAGVIFPAVCTRGVAKAVDFIPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 6  
S40770  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (HCV 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C/Species: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C/Accession: S40770; PC1285  
R/Okamoto, H.  
Submitted to the EMBL Data Library, March 1992  
A/Reference number: S40770  
A/Accession: S40770  
A/Molecule type: genomic RNA

A;Residues: 1-3011 <OKA>  
 A;Cross-references: EMBL:D10749; NID:G221586; PIDN:BA01582.1; PID:G221587  
 R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
 A;Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A;Reference number: F01284; MUID:91013116; PMID:2170712  
 A;Accession: F01285  
 A;Molecule type: genomic RNA  
 A;Residues: 1-513 <OK2>  
 A;Cross-references: GB:D00831; NID:G221511; PIDN:BA00705.1; PID:G221512  
 A;Experimental source: isolate HC-J1  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F;2-115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-389/Product: major envelope protein E #status predicted <ME>  
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F;1312-1317/Region: nucleotide-binding motif B  
 F;1316-1319/Region: DEXH motif  
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 Query Match 88.3%; Score 1403; DB 1; Length 3011;  
 Best Local Similarity 84.8%; Pred. No. 1.1e-112; Indels 0; Gaps 0;  
 Matches 257; Conservative 23; Mismatches 23;  
 Db 1 AGITVFYFVAAGLIRACMLVRKAAGHYVQMAFMKALATGYVYDHLTPLODWAHAG 60  
 904 ASLKVYFVFAVQGLIRFCALARKMGHYVQWAIITGLTGYVYNHLPRLADWAHAG 963  
 QY 61 LRLDAVAEPIVTSDEMEVKITTWGADTAAACDITISGIPVARSRRREILLGPADNFEQGM 120  
 Db 964 LRLDAVAEPIVTSDEMEVKITTWGADTAAACDITISGIPVARSRRREILLGPADNFEQGM 1023  
 QY 121 RLAPITAYSGOQTGGLGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 180  
 Db 1024 RLAPITAYSGOQTGGLGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 1083  
 QY 181 GAGSKTLAAGKPGPITOMTNTVDODLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 240  
 Db 1084 GAGRTITASPKGPVIQWNTVNDODLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 1143  
 QY 241 RRGSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 300  
 Db 1144 RRGSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 Db 1204 TMR 1206  
 RESULT 7  
 GNMVOC  
 genome polyprotein - hepatitis C virus (strain HCV-1)  
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 C;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C;Accession: A39166; P00403; P00404  
 R;Choo, Q.L.; Richman, R.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A;Title: Genetic organization and diversity of the hepatitis C virus.  
 A;Reference number: A39166; MUID:91172826; PMID:1848704  
 A;Accession: A39166  
 A;Molecule type: mRNA  
 A;Residues: 1-3011 <CHO>  
 A;Cross-references: GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874  
 R;Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Featherer, J.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992

A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A;Reference number: F00393; MUID:92268871; PMID:1136939  
 A;Accession: P00403  
 A;Molecule type: genomic RNA  
 A;Residues: 1577-1633 <CHA>  
 A;Cross-references: DDBJ:D10128  
 A;Experimental source: isolates E-b16  
 A;Accession: P00404  
 A;Status: Preliminary  
 A;Molecule type: genomic RNA  
 A;Residues: 1577-1633 <CH2>  
 A;Experimental source: isolates E-b17  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F;1-115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-389/Product: major envelope protein E #status predicted <ME>  
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F;1312-1317/Region: nucleotide-binding motif B  
 F;1316-1319/Region: DEXH motif  
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,224  
 Query Match 88.0%; Score 1398; DB 1; Length 3011;  
 Best Local Similarity 84.5%; Pred. No. 2.9e-112; Indels 0; Gaps 0;  
 Matches 256; Conservative 24; Mismatches 23;  
 Db 1 AGITVFYFVAAGLIRACMLVRKAAGHYVQMAFMKALATGYVYDHLTPLODWAHAG 60  
 904 ASLKVYFVFAVQGLIRFCALARKMGHYVQWAIITGLTGYVYNHLPRLADWAHAG 963  
 QY 61 LRLDAVAEPIVTSDEMEVKITTWGADTAAACDITISGIPVARSRRREILLGPADNFEQGM 120  
 Db 964 LRLDAVAEPIVTSDEMEVKITTWGADTAAACDITISGIPVARSRRREILLGPADNFEQGM 1023  
 QY 121 RLAPITAYSGOQTGGLGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 180  
 Db 1024 RLAPITAYSGOQTGGLGCIITSLTGDRKNQVEGEVQVSTATQSFATCNGVCWTFVH 1083  
 QY 181 GAGSKTLAAGKPGPITOMTNTVDODLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 240  
 Db 1084 GAGRTITASPKGPVIQWNTVNDODLVGMQAPPGARSMTPTCTGSSDLYLTRHADVIPIVR 1143  
 QY 241 RRGSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 300  
 Db 1144 RRGSRGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 Db 1204 TMR 1206  
 RESULT 8  
 GNMVCH  
 genome polyprotein - hepatitis C virus (strain H)  
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 A;Note: host Homo sapiens (man)  
 C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C;Accession: A36814; A41546  
 R;Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 submitted to GenBank, July 1992  
 A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: co  
 A;Reference number: A36814  
 A;Accession: A36814  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3011 <INC>

A:Cross-references: GB:J67463; NID:G9329737; PTDN:AAA45334.1; PID:G9329728  
R:Rincharnsape, G.; Zebade, U.S., Lee, D.H., Sugtanti, M.; Naeoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10282-10296, 1991  
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari  
A:Reference number: A41546; MUID:92052256; PMID:1658800  
A:Contents: annotation  
A>Note: neither amino acid nor nucleotide sequence is given  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
F:1.115/Product: capsid protein C #status predicted <CPC>  
F:1.16-191/Product: envelope protein M #status predicted <EPM>  
F:1.192-389/Product: major envelope protein E #status predicted <ME>  
F:3.90-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:7.703-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:10.7-1615/Product: hepatitisA virus #status predicted <NS3>  
F:12.30-1237/Region: nucleotide-binding motif A (P-loop)  
F:13.12-1317/Region: nucleotide-binding motif B  
F:13.16-1319/Region: DEXH motif  
F:1.616-1862/Product: nonstructural protein NS4A #status predicted <NS4A>  
F:1.663-2013/Product: nonstructural protein NS4B #status predicted <NS4B>  
F:2.014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:1.196, 209, 224, 305, 325, 417, 423, 430, 448, 476, 533, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 225

```

Query Match          87.2%; Score 1385; DB 1; Length 3011;
Best Local Similarity 83.8%; Pred. No. 3,9e-11;
Matches 254; Conservative 26; Mismatches 23; Indels 0; Gaps 0;

QY      1 AGITKVPYFVRAQGLIRACMLVRKAAGGHVYQMAFMKLAALTGYVDHLTELODMAHG 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      904 ASLLKVPFVFAVQGLIRICALRKLAGGHVQMAIKKLGALTGCVCVNHLPRLDMAHG 963
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      61 LRLDAVANEPIVPSMVEYKITTGADTACGDIISGLPVSRRRGEIILGGADNFEQGW 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      964 LRLDAVAEVPVFSMETKLIITWGADTACGDIINGLVSARRQEIILGGADNVSXGW 1023
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      121 RLAPITVYSOOTRELIGCIITSLTGRDKNVEGVQVSTATOSFLATCVNGCVTFEH 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1024 RLAPITVYAQOTRELIGCIITSLTGRDKNVEGVQIVSTATOTFLATCINGVCMTYIH 1083
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      181 GAGSKTLGPKGPIITQMYTNNVDOLVGMQAPGARSMTPCTGSSDLYLTRHADVIPVR 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1084 GAGTTLTASPGPVIQTITNNVDQDLVGMFAPQSGSLTPCTGSSDLYLTRHADVIPVR 1143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      241 RRGDSRGLSLSPRPVSYLKGSSGGPLLCPSGHANGVIFPAAVCTGVAKAUVFIVESET 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1144 RRGDSRGLSLSPRPSTYLKGSSGGPLLCPTEHANGVLPFAAVCTRGVAKAVDFIVENLET 1203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      301 TMR 303
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1204 TMR 1206
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9  
 JC5620  
 genome polypeptide - hepatitis C virus (isolate EUH480)  
 N:Contains: capsid protein C, envelope protein M, hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: JC5620  
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A>Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
 A:Reference number: JC5620; MUID:97366593; PMID:9223423  
 A:Accession: JC5620  
 A:Molecule type: mRNA  
 A:Residues: 1-3014 <CHA>  
 A:Cross-references: GB:Y13184  
 A:Experimental source: genotype 5a, which predominates in South Africa  
 A:Note: the translation of the nucleotide sequence is not complete in this paper  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleic acid binding; P-loop; polyprotein; serin  
 P:2-115/Product: capsid protein C #status predicted <CPC>

	F:116-191/Product: envelope protein M #status predicted <EPN>	
	F:182-389/Product: major envelope protein E #status predicted <ME>	
	F:384-408/Region: hypervariable #status predicted	
	F:350-730/Product: nonstructural protein NS1 #status predicted <NS1>	
	F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>	
	F:1008-1616/Product: hepacivirin #status predicted <NS3>	
	F:1231-1238/Region: nucleotide-binding motif A (P-loop)	
	F:1113-1318/Region: nucleotide-binding motif B	
	F:1117-1320/Region: DEAH motif	
	F:1617-1863/Product: nonstructural protein NS4a #status predicted <NA4>	
	F:1664-2014/Product: nonstructural protein NS4b #status predicted <NA5>	
	F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>	
	F:2210-2249/Region: interferon sensitivity determining #status predicted	
Query Match	77.7%; Score 1234; DB 1; Length 3014;	
Best Local Similarity	72.8%; Pred. No. 4,7e-98;	
Matches 219;	Conservative 43; Mismatches 39; Indels 0; Gaps 0;	
QY	3 ITRKVPFVRAOGIRACMLVRKAAAGHYVQAFMKLAALGTGVYDHLPLDQMAHAGR 62	
DQ	907 LTRKVPFLRRALRLRLCLAKHLVGRVQALHLGLTLGTYTHLAPMKDMASGR 966	
QY	63 DLAAVEPVIFSDMEVKIITWGDITACGDIISGLVPSARRGEIILGPADNFEQGMRL 122	
DQ	967 ELTVAETPIVFSAMETKRVITWGDITACGNIILAVLPVSARRGEIIFLGPADIKTSGMRL 1026	
QY	123 LAPITVASQGTRELCLCITSLTRPDKNOVEGVQVVSATOSPLATCNGVCWTFEHA 182	
DQ	1027 LAPITAAQGTREVLGAIVSLTRGDKNEBEGVQLSTATQTFELCLNGWMTLFEHA 1086	
QY	183 GSKTLGPKPIIQWYTNVDQDLVGVQADPGASMTPTCTGSSDLYLRHADVIPRRR 242	
DQ	1087 GSKTLGPKPKPIQWYTNVDXDLVGVSPSPGKSLRCTCGSADLYLRHADVIPARR 1146	
QY	243 GDSRGSILSRPVSYLKSGSGGPLCPDSGHAVCIIPAAVCTRVAQAQVDFIPVSEMTTM 302	
DQ	1147 GDTFRASILSRPPISTYLKSGSGGPIMGPSGHVGVFFAAVCTRVAALFEVVENLETTM 1206	
QY	303 R 303	
DQ	1207 R 1207	

RESULT 10  
J01303  
Genome polypeptide - hepatitis C virus (isolate HC-J6)  
N/contams: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
C/Accession: J01303  
R/Organism: H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iinaka, H.; Machida, A.; Miyakawa, Y.  
J. Gen. Virol. 72: 2697-2704, 1991  
A/Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
A/Reference number: J01303; MIMD:92044440; PMID:1658196  
A/Accession: J01303  
A/Molecule type: genomic RNA  
A/Residues: 1-303 <OK>  
A/Cross-references: GB:000944; NID:G221650; PIDN:BA00792.1; PID:G221651  
A/Experimental source: isolate HC-J6 from a Japanese individual  
C/Superfamily: hepatitis C virus genome polypeptide  
C/Keywords: ATP; glycoprotein; hydrolase; P-loop; polypeptide; serine proteinase; transmembrane  
F/2-115/Product: capsid protein C #status predicted <GPC>  
F/116-151/Product: envelope protein M #status predicted <EPM>  
F/199-389/Product: major envelope protein E #status predicted <NEE>  
F/390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F/734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1011-1619/Product: hepacivirin #status predicted <NS3>  
F/1136-1321/Region: nucleotide-binding motif B  
F/1130-1323/Region: DEH motif  
F/1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>  
F/1867-2011/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F.1130-1321/Region: nucleotide-binding motif B  
F.1130-1323/Region: DEXH motif  
F.11620-1866/Product: nonstructural protein NS4a #status predicted <NA>  
F.11867-2011/Product: nonstructural protein NS4b #status predicted <NA>  
F.2018-3033/Product: nonstructural protein NS5 #status predicted <NS>  
F.1196,209,533,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,235

Query Match  
Best Local Similarity 72.4%; Score 1150; DB 1; Length 3033;  
Matches 204; Conservative 47; Mismatches 52; Indels 0; Gaps 0;

QY 1 AGITKVVYEVFAOGLRACMLVRKAAGHYVQAFMKALITGYVYDHLTPLODMANAG 60  
Db ASLIRIPYFRAHALRVCITLVKHLGARYIQWLITIGRTGYIVDHLSPSTWAAQG 967  
QY 908  
Db 61 LRLDAVAEVIISDMEVKITITWGADTAACGDISLPASARGPILLGPADNFEQGW 120  
Db 968 LRLDAVAEVIISDMEVKITITWGADTAACGDISLPASARGPILLGPADNFEQGW 1027  
QY 121 RLAPAFVYSCQGRGLIGCIITSLTRGRKNQYGEVQVSTANQSHLATCVAGVCMYFH 180  
Db 1028 KLAAPAFVYSCQGRGLIGCIITSLTRGRKNQYGEVQVSTANQSHLATCVAGVCMYFH 1067  
QY 181 GAGSKTLAGEKGPITOMYTNVDDLVGKQAPPGASRMPTCTGSSDLYLVTHADYIPR 240  
Db 1088 GAGSKTLAGEKGPITOMYTNVDDLVGKQAPPGASRMPTCTGSSDLYLVTHADYIPR 1147  
QY 241 RRGSRGSLSPRPVSLTKSSGGPILCPGSHAVGIFRAVACPRGAKAVDPIPVESMET 300  
Db 1148 RKDRRCALSPRPVSLTKSSGGPILCPGSHAVGIFRAVACPRGAKAVDPIPVESMET 1207  
QY 301 TMR 303  
Db 1208 ATR 1210

RESULT 12  
T08841  
polyprotein - douroucouli hepatitis GB virus A  
C/Species: douroucouli hepatitis GB virus A  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #ext\_change 17-Nov-2000  
C/Accession: T08841  
R/Etker,U.J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montee, C.C.; Mushahwar, I.K.  
U. Gen. Virol. 79, 41-45, 1998  
A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
A/Reference number: Z16486; M01D:98120818; PMID:9460920  
A/Accession: T08841  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-3005 <ERK>  
A/Cross-References: EMBL:AF023425; NID:92828559; P1DN:AAC40502.1; PID:92828600  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: polyprotein

Query Match  
Best Local Similarity 25.0%; Score 397.5; DB 2; Length 3005;  
Matches 95; Conservative 46; Mismatches 119; Indels 13; Gaps 5;

QY 34 AFMKALITGYVYDHLTPLODMANAGLRDLAAVEVFIISDMEVKITITWGADTAACGDI 93  
Db 887 AFVRLERRRQVTLFQHCQGVSKXAAAILKDLGVALEPVSVTARDCVIARDAARTLACGOR 946  
QY 94 ISGLPISARGPILLG--PADNFEQGRRLAPITANSQOTRGLIGCIITSLTRGRKNQ 151  
Db 947 VEGLPVAVARGDVLGVFPSPVRLAPPGFVPTAPVV--MQRGLGFPVSVKTSMLGRDERE 1005  
QY 152 VEGEVQVSTATQSPFATCVNGCMTVFGAGSKTLAGEKGPITOMYTNVDDLVGKQAP 211  
Db 1006 HESIVYLGSTIRSRNGTCVNGMYTTFHGSNARTLAGVGCVNCRWMSPSDVAVYPLP 1065  
QY 212 PGRASMPCTCGSSDLYLVTHADYIPVRRGDSKSLSPRPVSLTKSSGGPILCPGSH 271  
Db 1066 SGASCPCEPKCGVQWCTRNL--DQALCHGRSKLVEIDLPEIISDFRSGSSPILCEBG 1123

QY 272 HAAGFRAAVCTRGV-----AKAVDFIPVES 297  
 DB 1124 HVGGMV-VSVLHRGVKTKGVKVPKMETLPRDS 1155

## RESULT 13

T08839  
 polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: T08839  
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <ERR>  
 A/Cross-references: EMBL:AF023424; NID:G2828597; PIDN:AC40501.1; PID:G2828598  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 21.5%; Score 341; DB 2; Length 2970;  
 Best Local Similarity 30.5%; Pred. No. 1e-20;  
 Matches 85; Conservative 46; Mismatches 96; Indels 52; Gaps 9;

QY 56 MAHAG-----LRDLAAVEPVIFSDMEVKIITGADTAAAGDIISGLPVSARRGRE 106  
 DB 891 YAHNGVTRTAEDLRQMGALBPVAVHPEDCAMVRAATLSGQGVHKGVPVARRGDE 950  
 QY 107 ILIGPADNFEQGMRL-----LAPITAYSQTRGLGCIITSLTGRDKNQVGEVQVVS 160  
 DB 951 VLIGVLNGV---WELDPGFVPTAPVYVH-HHKGFGGVKTSMTGMDTEHGVNVVLG 1005  
 QY 161 TATQSFATCVNGVGMVPHGAGSKTLGAPKPTQMYTNVDDLVGMQAPPARSMTPC 220  
 DB 1006 TSTRSGTCVNGVMTYTHSSNKTLLAQMGPVNSKMSASDVAVYPLPVAKCLPEPC 1065  
 QY 221 TCGSSDLYLVTRHADVIPRRRQDSRSGSLSS-----PRPVSYLKSSSGPPLCP 269  
 DB 1066 KCGQGVWVI-----RND--GALCHGTIGRTVELDLPALCDPFGSSGSPILCD 1112  
 QY 270 SGHAGVGFRAAVCTRG-----YAKAVDFIPVESMTT 301  
 DB 1113 EGHAVGML-ISVLRHGRSVTGIRYKPMETLPRALTHT 1150

## RESULT 14

B46642  
 DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chain  
 C/Species: Mus musculus (house mouse)  
 C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
 C/Accession: B46642  
 R/Miyazawa, H.; Irumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.  
 J. Biol. Chem. 268, 8111-8122, 1993  
 A/Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alpha  
 A/Reference number: A46642; MUID:93216786; PMID:8463324  
 A/Accession: B46642  
 A/Status: preliminary  
 A/Molecule type: mRNA; protein  
 A/Residues: 1-600 <MTY>  
 A/Cross-references: GB:D13546; NID:9303658; PIDN:BA02746.1; PID:9303659  
 A/Experimental source: FMA3 cells  
 A/Note: Sequence extracted from NCBI backbone (NCBI:129148, NCBI:P:129149)  
 C/Keywords: nucleotidyltransferase

Query Match 6.4%; Score 101; DB 2; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 0.85;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 15 LIRACMLVRKAAGHYQM-AFMKLAALT-----GTYYVDH-----TPLQDNA 57

DB 27 LAELCYLVQRTEDGAVGSELLAFCTSAKTCCTLVDTILNSFEVEVLNKLKSKAMHASKDSG 86  
 QY 58 HAGLDLAAVAPVIFSDMEVKIITWGDPTAACGDI--ISGLP-----VSARRGEI 107  
 DB 87 HAGTRDI-VSIQELLAESEETLLSSYTPSPGPKRVSTPTPLTKRSVAARSPRO- 144  
 QY 108 LIGPADNFEQGMRLIAPITAYSQTRGLGCIITSLTGRDKNQVGEVQVVSSTATQSEFL 167  
 DB 145 LSPFS-----FSPBATPSCK-----YISRNR-----GEVITTFGSAQ--- 178  
 QY 168 ATCVNGVCTVPHGAGSKTL--AGPKPTQMYTNVDDLVG 207  
 DB 179 -----GLSWSGRGGSGSVSLKVVGDPEPLTGSYKAMFOOLMG 215

## RESULT 15

G87392  
 conserved hypothetical protein CCL155 (imported) - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: G87392  
 R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I.  
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: G87392  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-353 <STO>  
 A/Cross-references: GB:AE005673; NID:913422473; PIDN:AAK23139.1; GSPDB:GN00148  
 C/Genetics:  
 A/Genes: CCL155

Query Match 6.3%; Score 99.5; DB 2; Length 353;  
 Best Local Similarity 21.8%; Pred. No. 0.59;  
 Matches 72; Conservative 36; Mismatches 114; Indels 109; Gaps 13;

QY 7 PYFVAQGLIRACMLVRKAA-----GGHYV-----QMAFMKLAALTGTYYVDH 50  
 DB 65 PLAVLAGLFAFSQGLARBSAIVMARASGSGYRIYGVAVPAVAVMLDALCGVTLAPRA 124  
 QY 51 TP-LQDMAHAGLRDLAAVEPVIFSDMEVKIITWGDPTAACGDIISGLPVSARRGEI 109  
 DB 125 DEPLADW-WRNTTPVAERKEPVPTFRAGADLVIGANASADRTLTGVTIFRRDSKGILV 183  
 QY 110 -----GPADNFEQGMRLIAPITAYSQTRGLGCIITSLTGRDKNQVGEVQVVSSTATQS 165  
 DB 184 EKVYEAARARYDGAATLLEQPT-----TRPADLSQAATA-- 219  
 QY 166 FLATCVNGVCTVPHGAGSKTLGAPKPTQMYTNVDDLVGMQAPPARSMTPCOTCCSS 225  
 DB 220 -----ATSWP-----TALRPDQVGLFDDSDMPAAAS----- 246  
 QY 226 DLYVTRHADVIPRRRQDSRSGSLSPRVSY-----LKGSSGSP-----LTCPSGHAVG 275  
 DB 247 -----ARRALENGG--SDRPESRYATHLQAFASPSVSLVMLLSAPVALA 290

Search completed: May 6, 2004, 09:37:15  
 Job time : 10.8681 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.32568 Seconds  
(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-10  
Sequence: 1 AGITKYPYFRAQGLIRACV.....RGVAKAVDPFVSEMTTMR 303

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	96.4	3010	1	POLG_HCVUT
2	1527	96.1	3010	1	POLG_HCVJA
3	1514	95.3	3010	1	POLG_HCVTW
4	1478	93.0	3010	1	POLG_HCVBK
5	1398	88.0	3011	1	POLG_HCVL
6	1385	87.2	3011	1	POLG_HCVL
7	1172	73.8	3033	1	POLG_HCVJ6
8	1150	72.4	3033	1	POLG_HCVJ6
9	101	6.4	600	1	DPO2_MOUSE
10	93	5.9	660	1	VST2_HEVBU
11	93	5.8	706	1	TRFE_HORSE
12	92.5	5.8	706	1	TRFE_HORSE
13	89.5	5.6	659	1	VST2_HEVME
14	88	5.5	3414	1	POLG_TBSVW
15	87	5.5	3414	1	POLG_TBSVW
16	87	5.5	3414	1	POLG_TBSVW
17	86.5	5.4	470	1	NRAM_IAMHM
18	86	5.4	434	1	NRAM_IAMHM
19	85	5.3	470	1	NRAM_IATRA
20	85	5.3	470	1	NRAM_IATRA
21	84.5	5.3	347	1	HELS_METMA
22	84.5	5.3	1705	1	MDHW_EUCGU
23	84	5.3	309	1	UCP2_MOUSE
24	84	5.3	339	1	UCP2_MOUSE
25	84	5.3	339	1	UCP2_MOUSE
26	83.5	5.3	538	1	DAC_ACTSP
27	83.5	5.3	854	1	PMW2_SCHPO
28	83	5.2	485	1	VST2_HEVRA
29	83	5.2	660	1	VST2_HEVRA
30	82.5	5.2	453	1	NRAM_IAMWL
31	82	5.2	309	1	UCP2_MOUSE
32	82	5.2	403	1	PGK_CHLMU
33	82	5.2	612	1	AMYG_ASFOR

34	81.5	5.1	398	1	TRNU_AGRIS	Q8u9ms agrobacteri
35	81.5	5.1	1022	1	CA26_CHICK	P15988 gallus gall
36	81	5.1	209	1	PAAD_PSEAE	O9ux08 pseudomonas
37	81	5.1	350	1	PE24_ARATH	O9z704 arabidopsis
38	81	5.1	470	1	NRAM_IAMUS	P03469 influenza a
39	81	5.1	730	1	HELS_METAC	O8t139 methanosarc
40	80	5.0	309	1	UCP2_HUMAN	P55851 homo sapien
41	80	5.0	326	1	PANE_RHILQ	Q987n5 rhizobium 1
42	80	5.0	961	1	ATCU_YERPE	O8zcat yerzinia pe
43	80	5.0	3491	1	ERY1_SACER	O03131 saccharopol
44	79.5	5.0	339	1	CRTB_RHCCA	P17056 rhodobacter
45	79.5	5.0	453	1	GAG_AVIMD	P06444 avian myelo

ALIGNMENTS

RESULT 1  
POLG\_HCVUT STANDARD; PRT; 3010 AA.  
AC Q00269;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polypeptide [contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4) (Hepatitisin)  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate HC-JT) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92295714; PubMed=1318627;  
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
RT Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
RT carrier: sequence variation within the same individual and among  
RT infected individuals.";  
RL Virus Res. 23:39-53(1992).  
CC -!- FUNCTION: The small proteins NS2, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polypeptide, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
CC (RNA) (N).  
CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein N and RNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL, D11168; BA01943.1; -;  
DR PIR, A45573; A45573.  
DR MEROPS, S29.001; -;  
DR MEROPS, U39.001; -;  
DR InterPro, IPR009003; Cys\_ser\_trypsin.  
DR InterPro, IPR001410; DEAD.  
DR InterPro, IPR002522; HCV\_capsid.





CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL; D90208; EAA14233.1; -;  
 DR PIR; A39253; GNMWCU.  
 DR HSSP; P26663; 1UXP.  
 DR MEROPS; U39.001; -;  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR00745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002668; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SMO0487; DEXDC1; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 136 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC...) (POTENTIAL)  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC...) (POTENTIAL)  
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC...) (POTENTIAL)  
 FT SEQUENCE 3010 AA; 327017 MW; AA93794F45DB185 CRC64;  
 -----  
 Query Match 96.1%; Score 1527; DB 1; Length 3010;  
 Best Local Similarity 94.4%; Pred. No. 4, 4e-124;  
 Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 AGTTKPVPRVRAQGLIRACMLYKRAAGHYVOMAFKALALGTYYDHLTPLODMANAG 60  
 DB 904 AGITRPVPRVRAQGLIRACMLYKRAAGHYVOMAFKALALGTYYDHLTPLODMANAG 963  
 QY 61 LRDIAVAEPVVFSPDMENVKIIITWADTAACGIIISGLPVSARGSEIILGPDNDEGQW 120  
 DB 964 LRDIAVAEPVVFSPDMENVKIIITWADTAACGIIISGLPVSARGSEIILGPDNDEGQW 1023  
 QY 121 RLAPITAVSQQTRGLLGLIITSITGRDNQVEGEVQVYSTATQSFATCVGVCMVYFH 180  
 DB 1024 RLAPITAVSQQTRGLLGLIITSITGRDNQVEGEVQVYSTATQSFATCVGVCMVYFH 1083  
 QY 181 GAGSKTLAEPKPIITOMYTNVODIVGQAPRGASMTPTCTGSSDLYLTHAVIPIR 240  
 DB 1084 GAGSKTLAEPKPIITOMYTNVODIVGQAPRGASMTPTCTGSSDLYLTHAVIPIR 1143  
 QY 241 RRGDSRGLSPRPVSYLKGSSGGLLCPSGHAYGIFPAAVCTRGAVAKADPIPVESMET 300  
 DB 1144 RRGDSRGLSPRPVSYLKGSSGGLLCPSGHAYGIFPAAVCTRGAVAKADPIPVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206  
 -----  
 RESULT 3  
 POLG\_HCVTM STANDARD; PRT; 3010 AA.  
 AC P29846;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide [contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21);  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepactivin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=31645;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230206; PubMed=1314449;  
 RA Chen P.T., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RT "The Taiwanese hepatitis C virus genome: sequence determination and  
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113 (1992).  
 CC -1- FUNCTION: The small proteins NS2a, NS2b, NS4a and NS4b are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipid-protein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M84754; -; NOT\_ANNOTATED\_CDS.

DR PIR; A40244; GNMVTM.

DR PDB; 1N64; 25-FEB-03.

DR PDB; 1NS3; 08-FEB-98.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

DR InterPro; IPR009003; Cys\_Ser\_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002518; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR002518; HCV\_NS4.

DR InterPro; IPR001490; HCV\_NS4a.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RdRp.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; peptidase\_C29.

DR InterPro; IPR007094; RNA\_pol\_DS\_PS.

DR Pfam; PF01543; HCV\_core; 1.

DR Pfam; PF01543; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR Pfam; PF01606; HCV\_NS1; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 789 789 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 327047 MW; AAD67D55CDEB215 CAC64;

Query Match 95.3%; Score 1514; DB 1; Length 3010;  
Best Local Similarity 93.1%; Pred. No. 5,9e-123;  
Matches 282; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 1 AGITKVPYFPAQGLIPACMLVRAAGHVVQVAFKALATGYVYDHLTPDQNAHAG 60  
DB 904 AGITRIPYFPAQGLIPACMLVRAAGHVVQVAFKALATGYVYDHLTPDQNAHAG 963  
QY 61 LRDLAVAEVPEVPSDMVEKIIITWADTAACGDIISGLPVASARRGRIILGPADNFGCGW 120  
DB 964 LRDLAVAEVPEVPSDMVEKIIITWADTAACGDIISGLPVASARRGRIILGPADNFGCGW 1023  
QY 121 RLAPITRYSQOTGLICITISLTERDKNOVEGEVQVSTATQSFATQNGVCTVPH 180  
DB 1024 RLAPITRYSQOTGLICITISLTERDKNOVEGEVQVSTATQSFATQNGVCTVPH 1083  
QY 181 GAGSKTLAAGPKPIPTQVYVNDQVLVGMQAPGARSMTPTCGSSDLYLTRADVIVR 240  
DB 1084 GAGSKTLAAGPKPIPTQVYVNDQVLVGMQAPGARSMTPTCGSSDLYLTRADVIVR 1143  
QY 241 RRGDSRGLSPRPVSYLKSGGGLICPSGHAVGIRAAVCTRGVAKADFLPVESMET 300  
DB 1144 RRGDSRGLSPRPVSYLKSGGGLICPSGHAVGIRAAVCTRGVAKADFLPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 4  
POLG\_HCVBK STANDARD; PRT; 3010 AA.  
ID POLG\_HCVBK  
AC P26663;  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate BK) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OC NCBI\_Taxid=11105;  
OX (1)  
RN SEQUENCE FROM N.A.  
RX MEDLINE=91140698; PubMed=1847440;  
RA Takamizawa A., Mori C., Fuke I., Tanabe S., Murakami S., Fujita J.,  
RA Onishi E., Andoh T., Yoshida I., Okayama H.;  
RT "Structure and organization of the hepatitis C virus genome isolated  
RT from human carriers.";  
RT J. Virol. 65:1105-1113 (1991).  
RN [2]  
RP SEQUENCE OF 1487-1500.  
RX MEDLINE=96235224; PubMed=8647104;  
RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetcky L.;  
RT "Non-structural protein 3 of hepatitis C virus inhibits





FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 88.0%; Score 1398; DB 1; Length 3011;  
 Best Local Similarity 84.5%; Pred. No. 7e-113;  
 Matches 256; Conservative 24; Mismatches 23; Indels 0; Gaps 0;

QY 1 AGIKKVEYFPAQGLIRACMLVRAAGHYQMAFMLALITGYVDHLPLDPAHAG 60  
 DB 904 ASLKVPEYFAVQGLRFLCALAKMIGHYQVWITLGLATGYVNHLPRLDMWANG 963  
 QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACDIIISGLPVASRRREILGPANFEGQGV 120  
 DB 964 LRLDAVAVEPIFSDMEVKIITWGADTAACDIIISGLPVASRRREILGPANFEGQGV 1023  
 QY 121 RLAPITAYSCQRTGLIGCIITSLTGDKNQVEGEVQVSTATQSPFATVNGVCTVPH 180  
 DB 1024 RLAPITAYSCQRTGLIGCIITSLTGDKNQVEGEVQVSTATQSPFATVNGVCTVPH 1083  
 QY 181 GAGSKTLAPGPKPTQVYTVNDQVLVGMQAPGASMTPTCCSSDLVLTFRHADVIVR 240  
 DB 1084 GAGSKTLAPGPKPTQVYTVNDQVLVGMQAPGASMTPTCCSSDLVLTFRHADVIVR 1143  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAKAVDFIVESMET 300  
 DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 6  
 POLG HCVA STANDARD; PRT; 3011 AA.

ID POLG HCVA  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/hellicase NS3 (P70) (Hepaticin)  
 DE (EC 3.4.21.38); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 CC Hepatitis C virus positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CC NCBI\_TaxID=11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchausti G., Zebadee S., Lee D.H.H., Sugtani M., Nasoff M.,  
 RA Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

RE MEDLINE=97313122; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain";  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RA MEDLINE=98154321; PubMed=9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding";  
 RL Structure 6:89-100(1998).  
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND NS5A.  
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: M67463; AAA45534.1; -  
 DR PIR: A36814; GNVCH.  
 DR PDB: 1HEI; 25-NOV-98.  
 DR PDB: 1AIV; 16-FEB-99.  
 DR PDB: 1AIR; 17-JUN-98.  
 DR MEROPS: S29.001; -.  
 DR TRANSFAC: T04155; -.  
 DR INTERPRO: IPR009003; Cys\_Ser\_trypsin.  
 DR INTERPRO: IPR001410; DEAD.  
 DR INTERPRO: IPR002522; HCV\_capsid.  
 DR INTERPRO: IPR002521; HCV\_core.  
 DR INTERPRO: IPR002519; HCV\_env.  
 DR INTERPRO: IPR002531; HCV\_NS1.  
 DR INTERPRO: IPR002518; HCV\_NS2.  
 DR INTERPRO: IPR000745; HCV\_NS4A.  
 DR INTERPRO: IPR001490; HCV\_NS4B.  
 DR INTERPRO: IPR002868; HCV\_NS5A.  
 DR INTERPRO: IPR002166; HCV\_NS5B.  
 DR INTERPRO: IPR001650; Helicase\_C.  
 DR INTERPRO: IPR004109; Peptidase\_C29.  
 DR INTERPRO: IPR007095; RNA\_pol\_DS\_PS.  
 DR INTERPRO: IPR007094; RNA\_pol\_PSVIR.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.



```

CC Hepcivlins.
CX NCBI_TaxId=11113;
RN [1]
RP SEQUENCE FROM N.A. MEDLINE=92044440; PubMed=1658196;
RX Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Iizuka H.,
RA Maehida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions."
RL J. Gen. Virol. 72:2697-2704(1991).
CC -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the precursor polypeptide, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N').
CC -I- SUBUNIT: The vitron of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-stb.ch/announce/ or send an email to license@isb-stb.ch).
-----
DR EMBL; D00944; BAA0792.1; -.
DR PIR; J01303; J01303.
DR HSSP; P27958; 1HRI.
DR MEROPS; S29.001; -.
DR InterPro; IPRO09003; Cys_Ser_trypsin.
DR InterPro; IPRO01410; DEAD.
DR InterPro; IPRO02522; HCV_capsid.
DR InterPro; IPRO02521; HCV_core.
DR InterPro; IPRO02519; HCV_env.
DR InterPro; IPRO02531; HCV_NS1.
DR InterPro; IPRO02518; HCV_NS2.
DR InterPro; IPRO00745; HCV_NS4B.
DR InterPro; IPRO01490; HCV_NS4A.
DR InterPro; IPRO02868; HCV_NS5A.
DR InterPro; IPRO02166; HCV_RdRp.
DR InterPro; IPRO01650; Helicase_C.
DR InterPro; IPRO04109; Peptidase_C29.
DR InterPro; IPRO07095; RNA_pol_DS_P6.
DR InterPro; IPRO07094; RNA_pol_PSvlt.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4A; 1.
DR Pfam; PF01001; HCV_NS4B; 1.
DR Pfam; PF01506; HCV_NS5A; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SMO0487; DEXC; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural
FT INT_MET 1
FT CHAIN 1 115
FT MATRIX 191
MATRIX PROTEIN (POTENTIAL).
CELLULAR ANTIOPEPTIDASE.
CAPSID PROTEIN C (POTENTIAL).
MATRIX PROTEIN (POTENTIAL).

```

Query	Best Local Similarity	73.8%; Score 1172; DB 1; Length 3033;
Matches	209; Conservative	44; Mismatches 48; Indels 0; Gaps 0;
QY 3	ITKVPYVRAQGLIRACMLVRKAGGYVMAFKLALTLGTIVYDHLTFLPDMAHAGRL	62
Db 910	LTRPYVYRAHALRLMCTVRHRLAGRGYVGMVLLALGRWTGYIYDHLTMSDMAANGRL	969
QY 63	DLANAVEPVYFSDMEVYKLTITMGADPTAACGGIITISGLPVSARKGRRLILGPDNFGCGWRL	122
Db 970	DLANAVEPIIFSPPEKKVITWGAETACGDIILGLPVSALGSEVLLGPADGYSKMSL	1029
QY 123	LAPYTAASQGTGRLGCIITSLGRDNGQVGEVGYVSTATQSFLATCVNAGCVMTVEHGA	182
Db 1030	LAPITAAAGTGRLLGLTIVVSMTRGRDTEBAGELQVLSITQSFLLGTTISGVLMTVYHGA	1089
QY 183	GSKTLAPKPGPIITOMYTNVDQDILVGCAPPGARSMPTCTCGSSDLIVYRHADYIPVRRR	242
Db 1090	GNNKLASRSRPVQMYMSABGDIVGWMPSPGRTSLBECTCGAADVLLVYTNNAVPIPARRR	1149
QY 243	GDSRGSITLSRPVSYLKSGSGPILCSGAVGIFRAVCTRGVAAKAVDPIPVESMETM	302
Db 1150	GDKRGALISRPISLTKSGSGGYVLCPRGAIVGFRAVCSRGVAAKSIDPIPVETIDIVT	1209
QY 303	R 303	
Db 1210	R 1210	

RESULT 8

POLG\_HCVU8 STANDARD; PRT; 3033 AA.

AC P26661;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 26-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (BC 3.4, 22.-); Protease/helicase NS3 (P70) (Hepadityrin)]







ID VST2 HEVPA STANDARD; PRT; 660 AA.  
 AC P33426;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Pakistan) (HEV).  
 OC Viruses; ssRNA positive-strand viruses; no DNA stage;  
 OC Hepatitis E-like viruses.  
 NC NCBI Taxid=33774;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92115700; Pubmed=1731327;  
 RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,  
 RA Malik I.A., Iqbal M., Purcell R.H.;  
 RT "Characterization of a prototype strain of hepatitis E virus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M80581; AAA45727.1; -  
 DR InterPro: IPR004261; SP2.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF03014; SP2; 1.  
 KM Signal.  
 FT SIGNAL 1 22 BY SIMILARITY  
 FT CHAIN 23 660 STRUCTURAL PROTEIN 2. CRC64;  
 SQ SEQUENCE 660 AA; 70980 MW; 80858C53CFB46FD3 CRC64;  
 Query March 5.9%; Score 93; DB 1; Length 660;  
 Best Local Similarity 19.0%; Pred. No. 2.3;  
 Matches 72; Conservative 49; Mismatches 117; Indels 140; Gaps 17;  
 QY 12 AGLIRACMLVRKAGHYVQAFMKALALTYVVDHLTPLODMAHAGRLDAVAVERV 71  
 DB 188 ARATIRYRPVVPNAVGVYAISIFWPQTITPTSV-----DNMSITSTVRLVQPG 239  
 QY 72 IFSMEVKITTWGADTAACDIIISGLVRSRGEILLGPAD--NFEQGMRL----- 123  
 DB 240 IASLVI-----PSRLHFRNQMREVSSTGVA 267  
 QY 124 -----APITAYSOQT-RGLIGCI-----ITSLTGRDKXQ----- 151  
 DB 268 EEEATSGVWMLCIHGSFVNSVTPTYGALGLDPALELEFRRLTGTNTVRYSYSTA 327  
 QY 152 -----VEGEVQVVSATQSELA-----TCVNGV-----CMTVTF----- 180  
 DB 328 RHRLRAGADGTAEITTAATREMKDLYFTSTNGVGEIGRIALITRLADLTIGLPTBL 387  
 QY 181 --GAG-----SKTLAGEPKG-PITOMYTNVDOLVGMQAPGASMTPTCGSSDLVY-- 230  
 DB 388 ISSAGGLLFYSRPVASNGEPTVXLYSVENA-----QODGIAIPHDIDGESRVVIOY 443  
 QY 231 -TRADVIPTARRRDSG-SLSRPVSYLK-----SSSGPELCPGHAVGIF 277  
 DB 444 DNGHEQDRPTSPSPASRPFVLRANDVLMSTLAAYDQSTGYSSSTGPVVY--SDSVTLV 501  
 QY 278 RAAVCTRGVAKAVDFIPV 295  
 DB 502 NVATGAQAVVASLDMTKV 519

ID TRFE HORSE STANDARD; PRT; 706 AA.  
 AC P27425;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Sero transferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal  
 DE binding globulin).  
 GN TF.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 NC NCBI Taxid=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92277958; Pubmed=8504171;  
 RA Carpenter M.A., Broad T.E.;  
 RT "The cDNA sequence of horse transferrin."  
 RL Biochim. Biophys. Acta 1173:230-232(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Extraembryonic tissue;  
 RA McDowell K.J., Adams M.H., Baker C.B.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.  
 CC -1- DOMAIN: Composed of two homologous domains.  
 CC -1- SIMILARITY: Belongs to the transferrin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M69020; AAA30958.1; -  
 DR EMBL; U21127; AAA63684.1; -  
 DR PIR; S33761; S33761.  
 DR HSP; P02787; IABE.  
 DR InterPro: IPR001156; Transferrin.  
 DR Pfam; PF00405; transferrin; 2.  
 DR PRINTS: PR00422; TRANSFERRIN.  
 DR SMART; SM00094; TR\_PFR\_2.  
 DR PROSITE; PS00205; TRANSFERRIN\_1; 2.  
 DR PROSITE; PS00206; TRANSFERRIN\_2; 2.  
 DR PROSITE; PS00207; TRANSFERRIN\_3; 2.  
 KM Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;  
 KW Signal.  
 FT SIGNAL 1 19 BY SIMILARITY  
 FT CHAIN 20 706 SEROTRANSFERRIN.  
 FT REPEAT 20 357 1.  
 FT REPEAT 358 706 2.  
 FT DISULFID 26 64 BY SIMILARITY.  
 FT DISULFID 36 55 BY SIMILARITY.  
 FT DISULFID 134 215 BY SIMILARITY.  
 FT DISULFID 174 190 BY SIMILARITY.  
 FT DISULFID 177 198 BY SIMILARITY.  
 FT DISULFID 187 200 BY SIMILARITY.  
 FT DISULFID 248 262 BY SIMILARITY.  
 FT DISULFID 360 623 BY SIMILARITY.  
 FT DISULFID 366 398 BY SIMILARITY.  
 FT DISULFID 376 389 BY SIMILARITY.  
 FT DISULFID 423 701 BY SIMILARITY.  
 FT DISULFID 441 664 BY SIMILARITY.  
 FT DISULFID 474 550 BY SIMILARITY.

RESULT 12  
 TRFE\_HORSE

```

FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;

```

```

Query Match 5.8%; Score 92.5; DB 1; Length 706;
Best Local Similarity 21.5%; Pred. No. 2.8;
Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

```

```

QY 33 MAEKLALNGTYV---YDHLTPLODVAHAGRLDVAVEPVSDEVKIITMGA----- 85
DB 321 LGFRIIPADMTWLYGLEYVT-----AIRNLREDIRPEVKD-ECKKVMKALGHH 371
QY 86 DTAACGD-IISGLPVASARRGR-----EILGFPADNFGQGWRL-----LAPITAY 129
DB 372 EKVCXDEWVSNGGNIACESAQSTEDCIATKVGEDAMSLDGFYIYAGKGLVPLAE 431
QY 130 SQCRGLGLGILTLTRDKNQVGEVQVYSTATQSLATCVAGVCTVTHGASKTLAG 189
DB 432 NYEIRSSGACVDPEESYH-----AAVAVKSSSDPDLT-----W-----NSLKG 470
QY 190 PKGPITOMYTNVDOLVGMQAPPGARSMTPTCTGSSDLVLTTHADVIPIRRRGDSRST 249
DB 471 KK-----SCHTGVDR-TAGWNI PMGL-----LYSEIHGCEPDKFRFGCAPGR 513
QY 250 LSRPVSYLKSSSGP-LIC-PSGHA-----VGIFFRAVCTRGVAKAVDFIPVESWE 299
DB 514 RNSTLCLNLGISASGPGRECEPNNHRYGYTGAFLVEKGDAV---FVKQATVE 566

```

```

RESULT 13
VST2 HEVME STANDARD; PRT; 659 AA.

```

```

AC Q03500;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor.
OS Hepatitis E virus (strain Mexico) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
ON NCBI_TaxID=31768;
RX MEDLINE=3079857; Pubmed=1448913;
RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,
RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
E virus (HEV)."
RL Virology 191:550-558 (1992).

```

```

CC -I- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74506; AAA5732.1;
DR PIR: B44212; B44212.
DR InterPro: IPR004261; SP2.
DR InterPro: IPR008975; Viral_cap_coat.
DR Pfam: PF03014; SP2; 1.
KM Signal.
FT CHAIN 1 22 BY SIMILARITY.
FT SIGNAL 23 659 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 659 AA; 70640 MW; C175B75BFD8FE2C CRC64;

```

```

Query Match 5.6%; Score 89.5; DB 1; Length 659;
Best Local Similarity 17.8%; Pred. No. 4.7;
Matches 67; Conservative 50; Mismatches 121; Indels 139; Gaps 14;

```

```

QY 12 AGLIRACLVYKAKAGHYVQAKEMKLAITGYVDHLTPLODVAHAGRLDVAVEPV 71
DB 188 ARATIRYRPLVNAVGAISISFWPQTTTPTSV-----DNNSITSTDRILVQPG 239
QY 72 IFSDMEVKIITWGAADTAACGDIISGLPVASRRREIILGPAD--NFGSQWRLL----- 123
DB 240 ISELVI-----APTANSQ-----QTRGLGICITISLGR----- 147
QY 124 -----APTANSQ-----QTRGLGICITISLGR----- 147
DB 268 EEEATSGVLMLCINHSFVNSYNTPTGTALGLDPALEEFNLTQWNTFRVSYSSTA 327
QY 148 --DKQVGEVGVVSTATQSFLLA-----TCVGVV-----CWTVHGASKTL 187
DB 328 KISARGADGTALITTAATRFKDLHFTGLNGVGVGEGIALTLNLADTLIGLPTLEI 387
QY 188 AGPKG-----PITOMYTNVDOLVGMQAPPGARSMTPTCTGSSDLV-- 230
DB 388 SSAGGQLFYSRPVVSANGEPYKLYTSVENA-----QDKGVAIPHIDILGDSRVVIQYD 443
QY 231 TTHADVIPIRRRGDSRG-SLSRPVSYLK-----GSSGGLPCSGHAYGIFR 278
DB 444 NQHEQDRPSPSPAPSPVSLRANVDLWLSLTAAYDOSYTGSSISGPVYI--SDSVTLVN 501
QY 279 AAVCTRGVAKAVDFIPV 295
DB 502 VATGAQAVARSLDMSKV 518

```

```

RESULT 14
POLG TBREV STANDARD; PRT; 3414 AA.
AC P14336; O88493;
ID POLG TBREV
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE (NS5)].
OS Tick-borne encephalitis virus (Western subtype) (TBEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
ON NCBI_TaxID=11088;
RX SEQUENCE FROM N.A., AND REVISIONS.

```





DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR001850; peptidase S7.  
 DR InterPro: IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVIR.  
 DR InterPro: IPR02877; Rnuv\_FtsJ.  
 DR Pfam: PF01003; Flavi\_capsid; 1.  
 DR Pfam: PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam: PF00869; Flavi\_glycoprot; 1.  
 DR Pfam: PF00948; Flavi\_helicase; 1.  
 DR Pfam: PF01004; Flavi\_M; 1.  
 DR Pfam: PF00948; Flavi\_NS1; 1.  
 DR Pfam: PF01005; Flavi\_NS2A; 1.  
 DR Pfam: PF01002; Flavi\_NS2B; 1.  
 DR Pfam: PF01350; Flavi\_NS4A; 1.  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR Pfam: PF01570; Flavi\_propep; 1.  
 DR Pfam: PF01578; FtsU; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR ProDom: PD001496; Flavi\_NS1; 1.  
 DR PROSITE: PS00690; DEAH ATP HELICASE; FALSE NEG.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;  
 KW ATP-binding; Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1  
 FT CHAIN 1 112  
 FT PROPEP 113 205  
 FT CHAIN 206 280  
 FT CHAIN 281 776  
 FT CHAIN 777 71128  
 FT CHAIN 71129 1358  
 FT CHAIN 1359 1489  
 FT CHAIN 1490 2110  
 FT CHAIN 2111 2259  
 FT CHAIN 2260 2510  
 FT CHAIN 2511 3412  
 FT NP\_BIND 1688 1695  
 FT SITE 1779 1782  
 FT TRANSMEM 101 112  
 FT TRANSMEM 247 259  
 FT TRANSMEM 266 280  
 FT TRANSMEM 738 751  
 FT TRANSMEM 738 751  
 FT DISULFID 340 396  
 FT DISULFID 340 396  
 FT DISULFID 354 385  
 FT DISULFID 372 401  
 FT DISULFID 466 570  
 FT DISULFID 587 618  
 FT CARBOHYD 144 144  
 FT CARBOHYD 434 434  
 FT CARBOHYD 861 861  
 FT CARBOHYD 983 983  
 FT CARBOHYD 999 999  
 FT CARBOHYD 1228 1228  
 FT CARBOHYD 2447 2447  
 FT CARBOHYD 2466 2466  
 FT CONFLICT 381 381  
 FT CONFLICT 850 850  
 SQ SEQUENCE 3412 AA; 377976 MW; 0F61CE6DCCDC5965 CRC64;  
 E -> S (IN REF. 3).  
 E -> D (IN REF. 3).  
 Query Match 5.5%; Score 87; DB 1; Length 3412;  
 Best Local Similarity 24.4%; Pred. No. 58; Mismatches 58; Indels 74; Gaps 11;  
 Matches 50; Conservative 23; WtMatches 58; Indels 74; Gaps 11;  
 QY 95 SGLFVSARRGRREILLGPADNFEQGGWELLAPITAYSQOTRGLGCIITSLTGRDKNQVEG 154  
 DB 1490 SGLFVSGGGRERDRFEEVDGV-YRIFSP-----GLLW-----G 1524  
 QY 155 EVQV-VSTAQSPLATCNGVCWTFVFGAG---SKTLAGEKXPITQMYTNVDQDIY----- 206  
 DB 1525 QRQVGVGYSKGVLT-----MMHVTGGAALSIDAVAGP-----YMAADVKEVDVCYGG 1573

QY 207 -----GMOA-----PGARSMPTCGSSDLYLVTRHADVIYVRRRGDSKSLSP 252  
 DB 1574 AMSLEKPKGETVQVNAHPFG-RHBEVHQCCPGSLLDLT-----GRRIGA 1617  
 QY 253 RPVSYLKSSGGPILCPBGAHVGF 277  
 DB 1618 VPIDAKGTSGSPILINSQGVVGLY 1642  
 Search completed: May 6, 2004, 09:31:49  
 Job time : 7.32568 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 28.339 Seconds

(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-10  
Perfect score: 1589  
Sequence: 1 AGITKVPYFPAQGLIACM.....RGVAKAVDFIPVSMETMR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_prodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1550	97.5	3010	12 Q9J3H7	Q9J3H7 hepatitis c
2	1545	97.2	3010	12 Q68826	Q68826 hepatitis c
3	1545	97.2	3010	12 P90191	P90191 hepatitis c
4	1544	97.2	3010	12 Q9DTE6	Q9DTE6 hepatitis c
5	1542	97.0	3010	12 Q9DTE4	Q9DTE4 hepatitis c
6	1541	97.0	3010	12 Q9DTE6	Q9DTE6 hepatitis c
7	1540	96.9	3010	12 P88603	P88603 hepatitis c
8	1539	96.9	3010	12 Q9J3H5	Q9J3H5 hepatitis c
9	1539	96.9	3010	12 Q807P3	Q807P3 hepatitis c
10	1537	96.7	3010	12 Q9J3F9	Q9J3F9 hepatitis c
11	1534	96.5	361	12 Q70815	Q70815 hepatitis c
12	1534	96.5	3008	12 Q9J3P4	Q9J3P4 hepatitis c
13	1534	96.5	3010	12 Q9J3H3	Q9J3H3 hepatitis c
14	1533	96.5	3010	12 Q9J3H2	Q9J3H2 hepatitis c
15	1533	96.5	3010	12 Q9J3I0	Q9J3I0 hepatitis c
16	1532	96.4	3010	12 Q9J3Y3	Q9J3Y3 hepatitis c

17	1532	96.4	3010	12 Q9J3H6	Q9J3H6 hepatitis c
18	1531	96.3	3010	12 Q9J3I7	Q9J3I7 hepatitis c
19	1531	96.3	3010	12 Q9J3I8	Q9J3I8 hepatitis c
20	1531	96.3	3010	12 Q9J3I6	Q9J3I6 hepatitis c
21	1531	96.3	3010	12 Q9J3I5	Q9J3I5 hepatitis c
22	1531	96.3	3010	12 Q9J3H6	Q9J3H6 hepatitis c
23	1530	96.3	3010	12 Q9J3H0	Q9J3H0 hepatitis c
24	1530	96.3	3010	12 Q9J3H9	Q9J3H9 hepatitis c
25	1530	96.3	3010	12 Q9J3H2	Q9J3H2 hepatitis c
26	1529	96.2	3013	12 Q9J3H4	Q9J3H4 hepatitis c
27	1528	96.2	1186	12 Q81755	Q81755 hepatitis c
28	1528	96.2	2284	12 Q81817	Q81817 hepatitis c
29	1528	96.2	3010	12 Q68788	Q68788 hepatitis c
30	1528	96.2	3010	12 P89966	P89966 hepatitis c
31	1528	96.2	3010	12 Q9DTE7	Q9DTE7 hepatitis c
32	1528	96.2	3010	12 Q9DTE0	Q9DTE0 hepatitis c
33	1528	96.2	3014	12 Q9J3H0	Q9J3H0 hepatitis c
34	1527	96.1	3010	12 Q9J3H2	Q9J3H2 hepatitis c
35	1527	96.1	3011	12 Q9DTE3	Q9DTE3 hepatitis c
36	1526	96.0	361	12 Q70818	Q70818 hepatitis c
37	1526	96.0	3010	12 Q9J3H6	Q9J3H6 hepatitis c
38	1526	96.0	3010	12 Q9J3H6	Q9J3H6 hepatitis c
39	1525	96.0	3010	12 Q9J3H8	Q9J3H8 hepatitis c
40	1524	95.9	3010	12 Q9J3I1	Q9J3I1 hepatitis c
41	1523	95.8	3010	12 Q9J3I5	Q9J3I5 hepatitis c
42	1523	95.8	3010	12 Q9DTE9	Q9DTE9 hepatitis c
43	1523	95.8	3010	12 Q9J3I4	Q9J3I4 hepatitis c
44	1523	95.8	3010	12 Q81541	Q81541 hepatitis c
45	1522	95.8	3010	12 Q9J3G3	Q9J3G3 hepatitis c

#### ALIGNMENTS

RESULT 1

ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.

AC Q9J3H7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_Taxid=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD15;

RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;

RT "Characteristics of hepatitis C viral genome associated with disease progression."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SUBMITTER: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; AF207756; AAF65946.1; -

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00804; P00804.

DR PIR; P80329; P80329.

DR HSSP; P26661; IUXP.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electon transporter activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO: 0016740; F:transferase activity; IEA.  
 DR GO: 0006518; P:electron transport; IEA.  
 DR GO: 0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO: 0006530; P:transcription; IEA.  
 DR GO: 0019079; P:viral genome replication; IEA.  
 DR GO: 0019087; P:viral transformation; IEA.  
 DR InterPro: IPR003003; Cys Ser trypsin.  
 DR InterPro: IPR003345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR005522; HCV\_capsid.  
 DR InterPro: IPR005521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002511; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01547; HCV\_core; 1.  
 DR Pfam: PF01533; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; HCV\_NS5b; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR Coated protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 KW polypeptide: RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SEQUENCE 3010 AA; 327365 MW; D86537317FFA106 CRC64;

Query Match 97.5%; Score 1550; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 2.2e-127;  
 Matches 290; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

DR 1 AGIRKVVYFYPAQGLIRACMLVRKAGHYVQMAFMKALALTGYVVDHLTPQDWAHAG 60  
 DB 904 AGIRKVVYFYPAQGLIRACMLVRKAGHYVQMAFMKALALTGYVVDHLTPQDWAHAG 963  
 DR 61 LRDVAVVEPYVFSDEMEVKIITWAGDTAACGDIISGLPVASARRREITLGPADNFEQGM 120  
 DB 964 LRDVAVVEPYVFSDEMEVKIITWAGDTAACGDIISGLPVASARRREITLGPADNFEQGM 1023  
 DR 121 RLAPITRAVSQQTGLIGCTITSLTGRDKNOVEGEVYSTATQSFATQVNCVMTVPH 180  
 DB 1024 RLAPITRAVSQQTGLIGCTITSLTGRDKNOVEGEVYSTATQSFATQVNCVMTVPH 1083  
 DR 181 GAGSKTLAAGPKPTQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRADYI PVR 240  
 DB 1084 GAGSKTLAAGPKPTQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRADYI PVR 1143  
 DR 241 RRGSRGSLSPRVSYTKSSGGPILCPSHANGITRAAVCTRGVAKADPIPVSMET 300  
 DB 1144 RRGSRGSLSPRVSYTKSSGGPILCPSHANGITRAAVCTRGVAKADPIPVSMET 1203  
 DR 301 TMR 303  
 DB 1204 TMR 1206

RESULT 2  
 G68826 PRELIMINARY; PRT; 3010 AA.  
 ID G68826  
 AC G68826  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-MOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 OS Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepciviruses  
 OK NCBI\_taxid=11103;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J3;  
 RA Cho M.J.;  
 RT "Molecular cloning of Hepatitis C virus genome from a single Japanese patient." (Sep-1991) to the EMBL/GenBank/DBJ databases.  
 RL Submitted.  
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: D14484; BAA03375.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: F00246; F00246.  
 DR PIR: F00804; F00804.  
 DR PIR: P50329; P50329.  
 DR HSSP: P26663; IUXP.  
 DR GO: 0016021; C:integral to membrane; IEA.  
 DR GO: 0019028; C:viral capsid; IEA.  
 DR GO: 0019031; C:viral envelope; IEA.  
 DR GO: 0005524; F:ATP binding; IEA.  
 DR GO: 0008026; F:ATP dependent helicase activity; IEA.  
 DR GO: 0016787; F:hydrolyase activity; IEA.  
 DR GO: 0003723; F:RNA binding; IEA.  
 DR GO: 0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO: 0008236; F:serine-type peptidase activity; IEA.  
 DR GO: 0005198; F:structural molecule activity; IEA.  
 DR GO: 0016740; F:transferase activity; IEA.  
 DR GO: 0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO: 0006530; P:transcription; IEA.  
 DR GO: 0019079; P:viral genome replication; IEA.  
 DR GO: 0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR005522; HCV\_capsid.  
 DR InterPro: IPR005521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002511; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01547; HCV\_core; 1.  
 DR Pfam: PF01533; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; HCV\_NS5b; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolyase; Nonstructural protein; Polypeptide; Transmembrane.  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SEQUENCE 3010 AA; 327150 MW; 7270F47984554FAD CRC64;

Query Match 97.2%; Score 1545; DB 12; Length 3010;  
 Best Local Similarity 96.4%; Pred. No. 6.1e-127;



Matches 292; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITVPYFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLPLODMAHAG 60  
 DB 904 AGITAVPEFVAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLPLODMAHAG 963  
 QY 61 LRDIAVAEVPYFSPMEVKIITWGADTAACDIIISGLPVSAARREIILGPDNFEQGW 120  
 DB 964 LRDIAVAEVPYFSPMEVKIITWGADTAACDIIISGLPVSAARREIILGPDNFEQGW 1023  
 QY 121 RLAPITAVSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMVTFH 180  
 DB 1024 RLAPITAVSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMVTFH 1083  
 QY 181 GAGSKTLAAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLYLTRHADVIPVR 240  
 DB 1084 GAGSKTLAAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLYLTRHADVIPVR 1143  
 QY 241 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 300  
 DB 1144 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

## RESULT 3

ID P90191 PRELIMINARY; PRT; 3010 AA.

AC P90191;  
 DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RX MEDLINE=95340824; Pubmed=7542279;  
 RA Emonoto N., Sakuma I., Asehana Y., Kurosaki M., Murakami T.,  
 RA Yamamoto C., Izumi N., Marumo F., Sato C.,  
 RT "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b.";  
 RL J. Clin. Invest. 96:224-230 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DB EMBL: D50482; BAA09073.1; -;  
 DB PIR: A61196; A61196.  
 DB PIR: P00254; P00254.  
 DB PIR: P00804; P00804.  
 DB PIR: P80329; P80329.  
 DB PIR: 1DXK: 12-JAN-01.  
 DB GO: GO:0016021; C: integral to membrane; IEA.  
 DB GO: GO:0019028; C: viral capsid; IEA.  
 DB GO: GO:0019031; C: viral envelope; IEA.  
 DB GO: GO:0005524; F: ATP binding; IEA.  
 DB GO: GO:0008265; F: ATP dependent helicase activity; IEA.  
 DB GO: GO:0003723; F: RNA binding; IEA.  
 DB GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DB GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DB GO: GO:0005198; F: structural molecule activity; IEA.  
 DB GO: GO:0016740; F: transferase activity; IEA.

DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_tyrp\_sln.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001480; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5a.  
 DR InterPro: IPR004109; Helicase\_C.  
 DR InterPro: IPR004109; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_P51r.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_RdRp; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; transferase; transmembrane.  
 FT CHAIN 1 191  
 FT CHAIN 192 383  
 FT CHAIN 384 809  
 FT CHAIN 810 1026  
 FT CHAIN 1027 1657  
 FT CHAIN 1658 1711  
 FT CHAIN 1712 1972  
 FT CHAIN 1973 2419  
 FT CHAIN 2420 3010  
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 97.2%; Score 1545; DB 12; Length 3010;

Best Local Similarity 96.0%; Pred. No. 6,1e-127; Matches 291; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGITVPYFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLPLODMAHAG 60  
 DB 904 AGITAVPEFVAQGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHLPLODMAHAG 963  
 QY 61 LRDIAVAEVPYFSPMEVKIITWGADTAACDIIISGLPVSAARREIILGPDNFEQGW 120  
 DB 964 LRDIAVAEVPYFSPMEVKIITWGADTAACDIIISGLPVSAARREIILGPDNFEQGW 1023  
 QY 121 RLAPITAVSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMVTFH 180  
 DB 1024 RLAPITAVSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCMVTFH 1083  
 QY 181 GAGSKTLAAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLYLTRHADVIPVR 240  
 DB 1084 GAGSKTLAAGPKGPIITQMTNTNDOLVGNQAPPGARSMTPTCGSSDLYLTRHADVIPVR 1143  
 QY 241 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 300  
 DB 1144 RRGDSRGSLSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 4  
 ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=HCV142;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL; AB049091; BAB18804.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; PS0329; PS0329.  
 DR HSSP; P26663; 1JXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008226; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006618; F:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_typsin.  
 DR InterPro: IPR000345; Cys\_Ser\_typsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_core.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_core.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4a.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5a.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01543; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4a; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00277; helicase\_C; 1.

DR Pfam; PF00998; Viral\_Rdrp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.  
 KM HYDROLASE; Nonstructural protein; Polypeptide.  
 SC SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;  
 Query Match 97.2%; Score 1544; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 7,4e-127;  
 Matches 290; Conservative 8; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 AGITKVPYFAAGLITACMLVRAAGHYVQAEMKALATGYVYDHLTPIDQMAHAG 60  
 DB 904 AGITRVFPYFAAGLITACMLVRAAGHYVQAEMKALATGYVYDHLTPIDMAHTG 963  
 QY 61 LRDIAVAEPIVFPEDMEKIIITWGAADTAAAGDIISGLPVASRRGREIILGPADNFEQGW 120  
 DB 964 LRDIAVAEPIVFPEDMETKIITWGAADTAAAGDIISGLPVASRRGREIILGPADNFEQGW 1023  
 QY 121 RLAPITAYSOQTGILGCIITSLTGSDKNQVEGEVYVSTATQSFATCNGVCWTFH 180  
 DB 1024 RLAPITAYSOQTGILGCIITSLTGSDKNQVEGEVYVSTATQSFATCNGVCWTFH 1083  
 QY 181 GAGSKTLAAGPKPIITOMTAVDODLVGMQAPPGARSMTPTCCSSDLYLTRADVTPVR 240  
 DB 1084 GAGSKTLAAGPKPIITOMTAVDODLVGMQAPPGARSMTPTCCSSDLYLTRADVTPVR 1143  
 QY 241 RRGDSRGLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVSEMET 300  
 DB 1144 RRGDSRGLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTRGVAKAVDFIVSEMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206  
 RESULT 5  
 ID Q9DTE4 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE4;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=HCV150;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL; AB049093; BAB18806.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00804; P00804.  
 DR PIR; PS0329; PS0329.  
 DR HSSP; P26663; 1JXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.

DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0006740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; Cys\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; Peptidase C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01539; HCV core; 1.  
 DR Pfam: PF01538; HCV NS1; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF002271; helicase C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR Pfam: PF018062; HCV NS1; 1.  
 DR Pfam: PF018062; HCV NS1; 1.  
 DR SMART: SM00467; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR Coat protein; Envelope protein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327324 MW; 3DB6CF249BD151C CRC64;

Query Match 97.0%; Score 1542; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 1,le-126;  
 Matches 290; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVFYFRAQGLIRACMLVRAAGHYVMAFMKALITGVYVDHLTPIDMAHAG 60  
 DB 904 AGITKVFYFRAQGLIRACMLVRAAGHYVMAFMKALITGVYVDHLTPIDMAHAGS 963  
 QY 61 LRLDAVAEPIVSDMEVKITWGADTAACDIIISGLPVASRRGREILLGPADNFGQGN 120  
 DB 964 LRLDAVAEPIVSDMEVKITWGADTAACDIIISGLPVASRRGREILLGPADNFGQGN 1023  
 QY 121 RLAPITAYSGOQRGLIGCIITSLTGRDKNOVEGEVYVSTATOSPLATGVNVCMTVH 180  
 DB 1024 RLAPITAYSGOQRGLIGCIITSLTGRDKNOVEGEVYVSTATOSPLATGVNVCMTVH 1083  
 QY 181 GAGSKTLAGEKPGITQYTYVVDLVGMQAPPGARSMTPCTCGSSDLVYTRHADV1PYR 240  
 DB 1084 GAGSKTLAGEKPGITQYTYVVDLVGMQAPPGARSMTPCTCGSSDLVYTRHADV1PYR 1143  
 QY 241 RRGDSRGSLSIPRVSTYLKSSSGGFLICSGHAGVIGTRAVVCTRGVAKAVDFPVESMET 300  
 DB 1144 RRGDSRGSLSIPRVSTYLKSSSGGFLICSGHAGVIGTRAVVCTRGVAKAVDFPVESMET 1203  
 QY 301 TMR 303  
 |||

DB 1204 TMR 1206  
 RESULT 6  
 ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxId=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV221;  
 RA Takashashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited";  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC EMBL: AB049101; BAB1814.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: P50329; P50329.  
 DR HSSP: P26663; LUXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0016787; F:hydrolase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; Cys\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; Peptidase C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.

DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR Pfam: PF0186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC1; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydroxylase; Nonstructural protein; Polyprotein;  
 KM RNA-directed RNA polymerase; Transferrase;  
 SQ SEQUENCE 3010 AA; 327108 MW; DE182D810EF78EE4 CRC64;

Query Match 97.0%; Score 1541; DB 12; Length 3010;  
 Best Local Similarity 96.0%; Pred. No. 1.4e-126;  
 Matches 291; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKYVYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMWANG 60  
 DB 904 AVLIKVFYFRAQGLIRACMLVRKAGHYQMAFMKLAALTGTYYVDHLTPLODMWANG 963

QY 61 LRLDLAAVEPVYFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPADNFEQGM 120  
 DB 964 LRLDLAAVEPVYFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPADNFEQGM 1023

QY 121 RLAPITAYVSOOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCNGVCWTVF 180  
 DB 1024 RLAPITAYVSOOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCNGVCWTVF 1083

QY 181 GAGSKTLAGPKGPIQTQMTNVDDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPIR 240  
 DB 1084 GAGSKTLAGPKGPIQTQMTNVDDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPIR 1143

QY 241 RRGDSRSLSIPRVSYLKGSSGGPILCPGSHVGIFFRAVCTRGVAKAVDFIPVESMET 300  
 DB 1144 RRGDSRSLSIPRVSYLKGSSGGPILCPGSHVGIFFRAVCTRGVAKAVDFIPVESMET 1203

QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 7  
 P88803 PRELIMINARY; PRT; 3010 AA.  
 AC P88803;  
 DT 01-MAY-1997 (TREMBLrel.03, Created)  
 DT 01-MAY-1997 (TREMBLrel.03, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel.25, Last annotation update)  
 DE Genome Polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 CC NCB1\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA Enomoto N.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA Enomoto N.; Sakuma I., Asahina Y., Kurosaki M., Murakami T.,  
 RA Yamamoto C., Izumi N., Martomo F., Sato C.;  
 RT "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b."  
 RU J. Clin. Invest. 96:224-230 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS.  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MNNA (BY SIMILARITY).  
 DR EMBL; D50484; BAA09075.1; -  
 DR PIR; A61196; A61196.

DR HSRP; P26663; INS3.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:structural molecule activity; IEA.  
 DR GO; GO:0005198; F:transferase activity; IEA.  
 DR GO; GO:0016740; F:transcription; IEA.  
 DR GO; GO:0006350; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_tyrpsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002519; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR001650; HCV\_RdRp.  
 DR InterPro; IPR004109; Helicase\_C.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01506; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF0186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC1; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 FT CHAIN 1  
 FT CHAIN 191  
 FT CHAIN 192 383  
 FT CHAIN 384 809  
 FT CHAIN 810 1026  
 FT CHAIN 1027 1657  
 FT CHAIN 1658 1711  
 FT CHAIN 1712 1972  
 FT CHAIN 1973 2419  
 FT CHAIN 2420 3010  
 SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEFPB8 CRC64;

Query Match 96.9%; Score 1540; DB 12; Length 3010;  
 Best Local Similarity 95.4%; Pred. No. 1.7e-126;  
 Matches 289; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGITKYVYFVRAQGLIRACMLVRKAAGHYQMAFMKLAALTGTYYVDHLTPLODMWANG 60  
 DB 904 AVLIKVFYFRAQGLIRACMLVRKAGHYQMAFMKLAALTGTYYVDHLTPLODMWANG 963

QY 61 LRLDLAAVEPVYFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPADNFEQGM 120  
 DB 964 LRLDLAAVEPVYFSMEVKIITWGADTAACGDIISGLPVSARRREIILGPADNFEQGM 1023

QY 121 RLAPITAYVSOOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCNGVCWTVF 180  
 DB 1024 RLAPITAYVSOOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCNGVCWTVF 1083

QY 181 GAGSKTLAGPKGPIQTQMTNVDDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPIR 240

```

Db      1084 GAGSKTLAAGKPIPTOMTNNVDQLVGMQAPPGARSILPTCTCGSSDLYLTRHADVIPIVR 1143
QY      241 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGAKAVDFIPVSEMET 300
Db      1144 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGAKAVDFIPVSEMET 1203
QY      301 TMR 303
Db      1204 TMR 1206

RESULT 8
ID      09J3H5      PRELIMINARY;      PRT; 3010 AA.
AC      09J3H5
DT      01-OCT-2000 (TrEMBLrel. 15, Created)
DT      01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Genome polypeptide.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MD17;
RA      Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT      "Characteristics of hepatitis C viral genome associated with disease
RL      progression.";
CC      Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA (BY SIMILARITY).
DR      EMBL; AF207758; AAF65948.1; -.
DR      PIR; A61196; A61196.
DR      PIR; P00246; P00246.
DR      PIR; P00254; P00254.
DR      PIR; P50329; P50329.
DR      HSP; P27958; IHE1.
DR      GO: GO:0016021; C: integral to membrane; IEA.
DR      GO: GO:0019028; C: viral capsid; IEA.
DR      GO: GO:0019031; C: viral envelope; IEA.
DR      GO: GO:0005524; F: ATP binding; IEA.
DR      GO: GO:0008026; F: ATP dependent helicase activity; IEA.
DR      GO: GO:0005489; F: electron transporter activity; IEA.
DR      GO: GO:0016787; F: hydrolase activity; IEA.
DR      GO: GO:0003723; F: RNA binding; IEA.
DR      GO: GO:0003668; F: RNA-directed RNA polymerase activity; IEA.
DR      GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR      GO: GO:0005198; F: structural molecule activity; IEA.
DR      GO: GO:0016740; F: transferase activity; IEA.
DR      GO: GO:0006118; P: electron transport; IEA.
DR      GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR      GO: GO:0006350; P: transcription; IEA.
DR      GO: GO:0019079; P: viral genome replication; IEA.
DR      GO: GO:0019087; P: viral transformation; IEA.
DR      InterPro: IPR009003; Cys Ser. trypsin.
DR      InterPro: IPR001410; DEAD.
DR      InterPro: IPR002523; HCV capsid.
DR      InterPro: IPR002521; HCV core.
DR      InterPro: IPR002519; HCV env.
DR      InterPro: IPR002518; HCV NS1.
DR      InterPro: IPR000745; HCV NS2.
DR      InterPro: IPR001490; HCV NS4a.
DR      InterPro: IPR002868; HCV NS5a.
DR      InterPro: IPR002166; HCV RdRp.
DR      InterPro: IPR001650; Helicase_C.
DR      InterPro: IPR004109; Peptidase_C29.
DR      InterPro: IPR007095; RNA_pol_DS_PS.

```

```

DR      InterPro: IPR007094; RNA_pol_PsVtr.
DR      Pfam; PF01543; HCV_capsid; 1.
DR      Pfam; PF01542; HCV_core; 1.
DR      Pfam; PF01539; HCV env; 1.
DR      Pfam; PF01560; HCV NS1; 1.
DR      Pfam; PF01538; HCV NS2; 1.
DR      Pfam; PF02307; HCV NS3; 1.
DR      Pfam; PF01006; HCV NS4a; 1.
DR      Pfam; PF01001; HCV NS4b; 1.
DR      Pfam; PF01506; HCV NS5a; 1.
DR      Pfam; PF00271; helicase_C; 1.
DR      Pfam; PF00998; Viral RdRp; 1.
DR      ProDom; PD186062; HCV NS1; 1.
DR      SMART; SM00487; DEXDC_1.
DR      PROSITE; PS00190; CYTOCHROME_C_1.
KW      ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW      Hydrolase; Nonstructural protein; Polypeptide;
KW      RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B93B7AA4B CRC64;

Query Match      96.9%; Score 1539; DB 12; Length 3010;
Best Local Similarity 95.7%; Pred. No. 2.1e-126;
Matches 290; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY      1 AGITKVPYFVRAQGIIRACMLVKKAAAGHYVMAFMKLAALTGTYYDHLTPLODMAHAG 60
Db      904 AGITRPYFVRAQGIIRACMLVKKAAAGHYVMAFMKLAALTGTYYDHLTPLODMAHAG 963
QY      61 LRDLAAVEPVFSPMEVKIITWGDVPAACGDIISGLPVASARSGEILIGPADNEGGGW 120
Db      964 LRDLAAVEPVFSPMEVKIITWGDVPAACGDIISGLPVASARSGEILIGPADNEGGGW 1023
QY      121 RLAPITVYSGQTRGLGIIISLGRPNQVEGVVSTATQSLFATCVNVCMTVPH 180
Db      1024 RLAPITVYSGQTRGLGIIISLGRPNQVEGVVSTATQSLFATCVNVCMTVPH 1083
QY      181 GAGSKTLAAGKPIPTOMTNNVDQLVGMQAPPGARSILPTCTCGSSDLYLTRHADVIPIVR 240
Db      1084 GAGSKTLAAGKPIPTOMTNNVDQLVGMQAPPGARSILPTCTCGSSDLYLTRHADVIPIVR 1143
QY      241 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGAKAVDFIPVSEMET 300
Db      1144 RRGDSRGSLLSPRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGAKAVDFIPVSEMET 1203
QY      301 TMR 303
Db      1204 TMR 1206

RESULT 9
ID      0807P3      PRELIMINARY;      PRT; 3010 AA.
AC      0807P3;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Polypeptide.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=WILE;
RA      MEDLINE=22047193; PubMed=12051758;
RA      Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
RA      Nio Y., Hosaka M., Miyazaki Y., Shimotohno K.;
RT      "Subgenomic replicon derived from a cell line infected with the
RT      hepatitis C virus.";
RL      Biochem. Biophys. Res. Commun. 293:993-999(2002).
DR      EMBL; AB080299; BACS4896.1; -.
DR      GO; GO:0019028; C: viral capsid; IEA.
DR      GO; GO:0019031; C: viral envelope; IEA.

```



DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral\_RBP; 1.  
 DR PRODOM; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9DB93B650C7 CRC64;

Query Match 96.7%; Score 1537; DB 12; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3.1e-126;  
 Matches 288; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGITKVFYFRAQGLIRACMLVRKAAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 60  
 DB 904 AGITRMYPYFRAQGLIRACMLVRKAAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 963  
 QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 120  
 DB 964 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 1023  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFH 180  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFH 1083  
 QY 181 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPCTGSSDLYLVRHADVIPVR 240  
 DB 1084 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPCTGSSDLYLVRHADVIPVR 1143  
 QY 241 RRGDSRGSLLSPRPVSYLKSGSGPGLCPSGHAGVIFRAAVCTRGVAKAVDFPVSME 300  
 DB 1144 RRGDSRGSLLSPRPVSYLKSGSGPGLCPSGHAGVIFRAAVCTRGVAKAVDFPVSME 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

## RESULT 11

ID 070815 PRELIMINARY; PRT; 361 AA.

AC 070815;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DE Polypeptide (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 RN NCBI  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98321154; PubMed=9656998;  
 RA Yamada K., Mori A., Seki M., Kimura J., Yuasa S., Matsunura Y.,  
 RA Miyamura T.;  
 RT "Critical point mutations for hepatitis C virus NS3 proteinase";  
 RL Virology 246:104-112(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,  
 RA Miyamura T.;  
 RT "Enzymatic characterization of purified NS3 serine proteinase of  
 RT hepatitis C virus expressed in Escherichia coli";  
 RL FEBS Lett. 378:37-42(1998).  
 DR EMBL; AB013620; BAA28498.1; -  
 DR HSSP; P27958; IHEI.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_Trypsin.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.  
 FT NON TER 1  
 FT NON TER 361  
 SQ SEQUENCE 361 AA; 38336 MW; 87DC310C76F4BCC3 CRC64;

Query Match 96.5%; Score 1534; DB 12; Length 361;  
 Best Local Similarity 95.0%; Pred. No. 3.5e-127;  
 Matches 288; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVFYFRAQGLIRACMLVRKAAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 60  
 DB 5 AGITRMYPYFRAQGLIRACMLVRKAAAGHYVMAFMKLAALGTGYVDHLTPLODMAHG 64  
 QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 120  
 DB 65 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVASARGREVLIGPADNFEQGM 124  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFH 180  
 DB 125 RLAPITAYSQOTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFH 184  
 QY 181 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPCTGSSDLYLVRHADVIPVR 240  
 DB 185 GAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGARSMTPCTGSSDLYLVRHADVIPVR 244  
 QY 241 RRGDSRGSLLSPRPVSYLKSGSGPGLCPSGHAGVIFRAAVCTRGVAKAVDFPVSME 300  
 DB 245 RRGDSRGSLLSPRPVSYLKSGSGPGLCPSGHAGVIFRAAVCTRGVAKAVDFPVSME 304  
 QY 301 TMR 303  
 DB 305 TMR 307

## RESULT 12

ID 093F4 PRELIMINARY; PRT; 3008 AA.

AC 093F4;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 GN MD34.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 RN NCBI  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD34;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 RT progression";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 DR EMBL; AF208024; AAF61205.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P80329; P80329.  
 DR HSSP; P26663; IUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008025; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR000345; CysC\_heme\_B5.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002521; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002531; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR ARF-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolyase; Nonstructural protein; Transferase; Transmembrane.  
 KW RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SQ SEQUENCE 3008 AA; 326834 MW; 99AE05E14C3109F4 CRC64;

Query Match 96.5%; Score 1534; DB 12; Length 3008;  
 Best Local Similarity 95.7%; Pred. No. 5,7e-126;  
 Matches 290; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 AGITKVYFFRAOGLIACMLVKKAKAGHYVQMAFMKLAITGTYYVDHTPLQDPAHAG 60  
 DB 902 ASIIXKVYFVRAOGLIACMLARKVAGHYVQMAFMKLAITGTYYVDHTPLQDPAHAG 961  
 QY 61 LRDLAVAVEPFIPEDEMEVKIITWGADTAACDIIISGLPVASRRGREILLGPADNFEQGM 120  
 DB 962 LRDLAVAVEPFIPEDEMEVKIITWGADTAACDIIISGLPVASRRGREILLGPADNFEQGM 1021  
 QY 121 RLAPITAYSQQRTGLIGCIITSLTGDKNQVEGEVQVSTAFQSLATVNGVCTVHR 180  
 DB 1022 RLAPITAYSQQRTGLIGCIITSLTGDKNQVEGEVQVSTAFQSLATVNGVCTVHR 1081  
 QY 181 GAGSKTLAGKPGITQMTYTVNDQVLVQMAFGARSWPTCCSSSDIYVTRADVIPIVR 240  
 DB 1082 GAGSKTLAGKPGITQMTYTVNDQVLVQMAFGARSWPTCCSSSDIYVTRADVIPIVR 1141  
 QY 241 RRGDSRSGSLSPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 1142 RRGDSRSGSLSPRVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 1201  
 QY 301 TMR 303  
 DB 1202 TMR 1204

ID Q9J3H3 PRELIMINARY; FRT; 3010 AA.  
 AC Q9J3H3  
 DT 01-OCT-2000 (TRENDELrel. 15, Created)  
 DT 01-OCT-2000 (TRENDELrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TRENDELrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepciviruses.  
 OX NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD19.  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Saito C.;  
 RT Characteristics of hepatitis C viral genome associated with disease  
 RT progression.  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL, AF207760; AAE65950.1; -.  
 DR PIR; A61196; A61196.  
 DR HSRP; P26663; IUXP.  
 DR GO:0016021; C:Integral to membrane; IEA.  
 DR GO:0019028; C:Viral capsid; IEA.  
 DR GO:0019031; C:Viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR00345; CysC\_heme\_B5.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002521; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002531; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.

RESULT 13  
 Q9J3H3



SEQUENCE 3010 AA; 327234 MW; 44C3467649CB8DD CRC64;  
Query Match 96.5%; Score 1534; DB 12; Length 3010;  
Best Local Similarity 94.4%; Pred. No. 5,7e-126;  
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;  
QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMLALTGTYYVDHLTPLOMAHAG 60  
DB 904 AGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMLALTGTYYVDHLTPLOMAHAG 963  
QY 61 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRGREILLGPADNFEQGW 120  
DB 964 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRGREILLGPADNFEQGW 1023  
QY 121 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFATCVNGVCTVPH 180  
DB 1024 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFATCVNGVCTVPH 1083  
QY 181 GAGSKTIAGPKGPIITOMYTNVDDLVGMQAPGARSWTPTCGSSDLYLTRHADYIPVR 240  
DB 1084 GAGSKTIAGPKGPIITOMYTNVDDLVGMQAPGARSWTPTCGSSDLYLTRHADYIPVR 1143  
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206  
RESULT 14  
09J3H2 PRELIMINARY; PRT: 3010 AA.  
AC Q9J3H2; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MD20;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.,  
RT "Characteristics of hepatitis C viral genome associated with disease  
RL progression";  
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).  
DR EMBL, AF207761, AA65951.1, -.  
DR PIR, A61196, A61196.  
DR PIR, P00246, P00246.  
DR HSSP, P26663, INS3.  
DR GO:0016021; C:integral to membrane; IEA.  
DR GO:0019028; C:viral capsid; IEA.  
DR GO:0019031; C:viral envelope; IEA.  
DR GO:0005324; F:ATP binding; IEA.  
DR GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO:0005489; F:electron transporter activity; IEA.  
DR GO:0003723; F:RNA binding; IEA.  
DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO:0005198; F:structural molecule activity; IEA.  
DR GO:0016740; F:transferase activity; IEA.  
DR GO:0006118; P:electron transport; IEA.  
DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO:0006350; P:transcription; IEA.

DR GO:0019079; P:viral genome replication; IEA.  
DR GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR009003; Cys\_Ser\_lypsin.  
DR InterPro; IPR000345; CysC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV capsid.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR002518; HCV NS2.  
DR InterPro; IPR000745; HCV NS4a.  
DR InterPro; IPR001490; HCV NS4b.  
DR InterPro; IPR002668; HCV NS4b.  
DR InterPro; IPR002166; HCV NS4b.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA\_pol\_DS\_Ps.  
DR InterPro; IPR007094; RNA\_pol\_PsVtr.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
DR Pfam; PF01538; HCV NS2; 1.  
DR Pfam; PF02907; HCV NS3; 1.  
DR Pfam; PF01006; HCV NS4a; 1.  
DR Pfam; PF01001; HCV NS4b; 1.  
DR Pfam; PF01506; HCV NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR ProDom; PD186062; HCV NS1; 1.  
DR SMART; SMC0487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; 1.  
DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW polyprotein; RNA-directed RNA polymerase; transferase; Transmembrane.  
SQ SEQUENCE 3010 AA; 326763 MW; 1A48BE48E51440D0 CRC64;  
Query Match 96.5%; Score 1533; DB 12; Length 3010;  
Best Local Similarity 95.0%; Pred. No. 7e-126;  
Matches 286; Conservative 9; Mismatches 6; Indels 0; Gaps 0;  
QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMLALTGTYYVDHLTPLOMAHAG 60  
DB 904 AGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMLALTGTYYVDHLTPLOMAHAG 963  
QY 61 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRGREILLGPADNFEQGW 120  
DB 964 LRDIAVAVEPVFSDMEVKIITWGADTAACGDIISGLPVASRGREILLGPADNFEQGW 1023  
QY 121 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFATCVNGVCTVPH 180  
DB 1024 RLAPITAYSQQTRGLIGCIITSLTGRDNQVGEVQVSTATQSPFATCVNGVCTVPH 1083  
QY 181 GAGSKTIAGPKGPIITOMYTNVDDLVGMQAPGARSWTPTCGSSDLYLTRHADYIPVR 240  
DB 1084 GAGSKTIAGPKGPIITOMYTNVDDLVGMQAPGARSWTPTCGSSDLYLTRHADYIPVR 1143  
QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206  
RESULT 15  
09J3H2 PRELIMINARY; PRT: 3010 AA.  
AC Q9J3H2; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polyprotein.

OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD12;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBMITTER: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL; AF207753; AAF65943.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; PS0329; PS0329.  
 DR HSBP; P26663; IUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:Electron transporter activity; IEA.  
 DR GO; GO:0016787; F:Hydrolase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; F:electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:Viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR003003; Cys Ser tyrosin.  
 DR InterPro; IPR000345; CysC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5a.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydroxylase; Nonstructural protein; Polypeptide;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326692 MW; 074098DB305AF1A9 CRC64;

Query Match

96.5%; Score 1533; DB 12; Length 3010;

Best Local Similarity 95.4%; Pred. No. 7e-126;  
 Matches 289; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 AGITKVPYFVPAQGLIRACMLVRKAGGHVYQMAFMKLAALTGYVDHLTPLODWAHAG 60  
 DB 904 AGITRVFVFAQGLIRACMLVRKAGGHVYQMAFMKLAALTGYVDHLTPLRGWAHTG 963  
 QY 61 LRDIAVAEPIFSDMEVKIITWGAADPAACGDIISGLPVASRRREILLGPADNFEQGN 120  
 DB 964 LRDIAVAEPIFSDMEVKIITWGAADPAACGDIISGLPVASRRREILLGPADNFEQGN 1023  
 QY 121 RLAPITVYSGQTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVFH 180  
 DB 1024 RLAPITVYSGQTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVFH 1083  
 QY 181 GAGSKTLAGPKPTTQMTTNDQDLVGMQAPPGARSMTPCTCGSSDLYLTRHADVTPVR 240  
 DB 1084 GAGSKTLAGPKPTTQMTTNDQDLVGMQAPPGARSMTPCTCGSSDLYLTRHADVTPVR 1143  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAKAVDFIPVESMET 300  
 DB 1144 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAKAVDFIPVESMET 1203  
 QY 301 TWR 303  
 DB 1204 TWR 1206

Search completed: May 6, 2004, 09:35:45  
 Job time : 29.339 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 11.8923 seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-10  
Perfect score: 1589  
Sequence: 1 AGITKVPYFVRAGGLIRACM.....RGVAKAVDPFVPSMETTMR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgnt2\_6/ptodata/2/1aa/54\_COMB.pep:\*  
2: /cgnt2\_6/ptodata/2/1aa/58\_COMB.pep:\*  
3: /cgnt2\_6/ptodata/2/1aa/62\_COMB.pep:\*  
4: /cgnt2\_6/ptodata/2/1aa/68\_COMB.pep:\*  
5: /cgnt2\_6/ptodata/2/1aa/72\_COMB.pep:\*  
6: /cgnt2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1530	96.3	2201 4 US-09-539-601-6	Sequence 6, Appl1
2	1530	96.3	2201 4 US-09-539-601-15	Sequence 15, Appl1
3	1530	96.3	3010 4 US-09-539-601-3	Sequence 3, Appl1
4	1530	96.3	3010 4 US-09-539-601-21	Sequence 21, Appl1
5	1530	96.3	3010 4 US-09-539-601-27	Sequence 27, Appl1
6	1527	96.1	1692 3 US-09-263-933-4	Sequence 4, Appl1
7	1527	96.1	1692 3 US-09-263-933-2	Sequence 2, Appl1
8	1527	96.1	2307 3 US-09-263-933-11	Sequence 11, Appl1
9	1527	96.1	1692 3 US-09-263-933-9	Sequence 9, Appl1
10	1524	95.9	1692 4 US-09-919-901-11	Sequence 11, Appl1
11	1524	95.9	2307 3 US-09-919-901-9	Sequence 9, Appl1
12	1524	95.9	2307 3 US-09-919-901-16	Sequence 16, Appl1
13	1524	95.9	2307 3 US-09-919-901-12	Sequence 12, Appl1
14	1523	95.8	3010 4 US-09-539-601-33	Sequence 33, Appl1
15	1515	95.3	1692 3 US-09-263-933-18	Sequence 18, Appl1
16	1515	95.3	1692 3 US-09-263-933-15	Sequence 15, Appl1
17	1515	95.3	2307 3 US-09-263-933-11	Sequence 11, Appl1
18	1515	95.3	2307 3 US-09-263-933-9	Sequence 9, Appl1
19	1504	94.7	3010 3 US-09-014-816-3	Sequence 3, Appl1
20	1478	93.0	2013 2 US-08-324-977-12	Sequence 12, Appl1
21	1478	93.0	2013 2 US-08-384-616-12	Sequence 12, Appl1
22	1478	93.0	2013 2 US-08-904-686A-12	Sequence 12, Appl1
23	1478	93.0	2013 3 US-09-315-850-12	Sequence 12, Appl1
24	1478	93.0	2201 4 US-08-952-981A-2	Sequence 2, Appl1
25	1478	93.0	2620 1 US-08-324-977-32	Sequence 32, Appl1
26	1478	93.0	2620 1 US-08-384-616-32	Sequence 32, Appl1
27	1478	93.0	2620 2 US-08-904-686A-32	Sequence 32, Appl1

28	1478	93.0	2620 3 US-09-315-850-32	Sequence 32, Appl1
29	1478	93.0	2621 1 US-08-324-977-36	Sequence 36, Appl1
30	1478	93.0	2621 2 US-08-384-616-36	Sequence 36, Appl1
31	1478	93.0	2621 2 US-08-904-686A-36	Sequence 36, Appl1
32	1478	93.0	2621 2 US-09-315-850-36	Sequence 36, Appl1
33	1478	93.0	3010 1 US-08-324-977-2	Sequence 2, Appl1
34	1478	93.0	3010 1 US-08-324-977-14	Sequence 14, Appl1
35	1478	93.0	3010 2 US-08-384-616-2	Sequence 2, Appl1
36	1478	93.0	3010 2 US-08-384-616-14	Sequence 14, Appl1
37	1478	93.0	3010 2 US-08-904-686A-2	Sequence 2, Appl1
38	1478	93.0	3010 2 US-08-904-686A-14	Sequence 14, Appl1
39	1478	93.0	3010 2 US-09-315-850-2	Sequence 2, Appl1
40	1478	93.0	3010 3 US-09-315-850-14	Sequence 14, Appl1
41	1405	88.4	3012 3 US-08-811-566-2	Sequence 2, Appl1
42	1405	88.4	3012 4 US-09-034-756-2	Sequence 2, Appl1
43	1402	88.2	1648 1 US-08-188-281B-12	Sequence 12, Appl1
44	1402	88.2	1648 1 PCT-US94-07280-12	Sequence 12, Appl1
45	1402	88.2	1648 5 PCT-US95-01087-12	Sequence 12, Appl1

ALIGNMENTS

RESULT 1  
US-09-539-601-6  
Sequence 6, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartenschlager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 2201  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match 96.3%; Score 1530; DB 4;  
Best Local Similarity 95.0%; Pred. No. 2.5e-144;  
Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY	1	AGITKVPYFVRAGGLIRACMIVRKAGHYVQMAFMKLAALGTGVYDHLTPLODMAHAG	60
DB	95	AGITKVPYFVRAGGLIRACMIVRKAGHYVQMAFMKLAALGTGVYDHLTPLODMAHAG	154
QY	61	LEDLAFAVEPVFESDMETKVIITWADTAACGDIISGLFVSARSGEIIIGPADNFEQGW	120
DB	155	LEDLAFAVEPVFESDMETKVIITWADTAACGDIISGLFVSARSGEIIIGPADNFEQGW	214
QY	121	RLAATITVSOOTRGLGCIITSLTGRDNQVEGVVVSATOSFLATCVNGVCMTFPH	180
DB	215	RLAATITVSOOTRGLGCIITSLTGRDNQVEGVVVSATOSFLATCVNGVCMTFPH	274
QY	181	GAGSKTLAPKPKPIITONTVNDQDVLVQAPPGASMTPTCTGSSDYLVTIRHADIVFR	240
DB	275	GAGSKTLAPKPKPIITONTVNDQDVLVQAPPGASMTPTCTGSSDYLVTIRHADIVFR	334
QY	241	RQCDRSGSLSRPVSYLKSGSGGGLCPSGHANGIFPAAVCTRGVAKAVDPFVPSMET	300
DB	335	RQCDRSGSLSRPVSYLKSGSGGGLCPSGHANGIFPAAVCTRGVAKAVDPFVPSMET	394
QY	301	TMR 303	
DB	395	TMR 397	

RESULT 2

US-09-539-601-15  
 ; Sequence 15, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 15  
 ; LENGTH: 2201  
 ; TYPE: PRN  
 ; ORGANISM: Hepatitis C virus  
 ; US-09-539-601-15

Query Match 96.3%; Score 1530; DB 4; Length 2201;  
 Best Local Similarity 95.0%; Pred. No. 2.5e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 DB 95 AGITKVPYFVRAHGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 154  
 QY 61 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFEQGM 120  
 DB 155 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFEQGM 214  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQSLATCVNGVCWTVFH 180  
 DB 215 RLAPITAYSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQSLATCVNGVCWTVFH 274  
 QY 181 GAGSKTAGPKPITQMTNVDOLVGMQAPPGARSPTCTCGSSDLVYTRHADVIPIVR 240  
 DB 275 GAGSKTAGPKPITQMTNVDOLVGMQAPPGARSPTCTCGSSDLVYTRHADVIPIVR 334  
 QY 241 RRGDSRGSLSPPRVSYLKSSGGGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 335 RRGDSRGSLSPPRVSYLKSSGGGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 394  
 QY 301 TMR 303  
 DB 395 TMR 397

## RESULT 3

US-09-539-601-3  
 ; Sequence 3, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 3010  
 ; TYPE: PRN  
 ; ORGANISM: Hepatitis C virus  
 ; US-09-539-601-3

Query Match 96.3%; Score 1530; DB 4; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3.9e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 60

DB 904 AGITKVPYFVRAHGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 QY 61 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFEQGM 120  
 DB 964 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFEQGM 1023  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQSLATCVNGVCWTVFH 180  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQSLATCVNGVCWTVFH 1083  
 QY 181 GAGSKTAGPKPITQMTNVDOLVGMQAPPGARSPTCTCGSSDLVYTRHADVIPIVR 240  
 DB 1084 GAGSKTAGPKPITQMTNVDOLVGMQAPPGARSPTCTCGSSDLVYTRHADVIPIVR 1143  
 QY 241 RRGDSRGSLSPPRVSYLKSSGGGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 1144 RRGDSRGSLSPPRVSYLKSSGGGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

## RESULT 4

US-09-539-601-21  
 ; Sequence 21, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 21  
 ; LENGTH: 3010  
 ; TYPE: PRN  
 ; ORGANISM: Hepatitis C virus  
 ; US-09-539-601-21

Query Match 96.3%; Score 1530; DB 4; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 3.9e-144;  
 Matches 288; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 DB 904 AGITKVPYFVRAHGLIRACMLVRKAGHYQVAFMKLAALTGTYYVDHLTPLODMAHAG 963  
 QY 61 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFEQGM 120  
 DB 964 LRDIAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFEQGM 1023  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQSLATCVNGVCWTVFH 180  
 DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQSLATCVNGVCWTVFH 1083  
 QY 181 GAGSKTAGPKPITQMTNVDOLVGMQAPPGARSPTCTCGSSDLVYTRHADVIPIVR 240  
 DB 1084 GAGSKTAGPKPITQMTNVDOLVGMQAPPGARSPTCTCGSSDLVYTRHADVIPIVR 1143  
 QY 241 RRGDSRGSLSPPRVSYLKSSGGGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 1144 RRGDSRGSLSPPRVSYLKSSGGGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMET 1203  
 QY 301 TMR 303  
 DB 1204 TMR 1206

RESULT 5  
US-09-539-601-27  
; Sequence 27, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlagel, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-27

Query Match 96.3%; Score 1530; DB 4; Length 3010;  
Best Local Similarity 95.0%; Pred. No. 3.5e-144;  
Matches 286; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
DB 904 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYYVDHLTPLODMAHAG 963  
QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 120  
DB 964 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 1023  
QY 121 RLAPITAYSOOTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 180  
DB 1024 RLAPITAYSOOTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 1083  
QY 181 GAGSKTLAPKPGPITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 240  
DB 1084 GAGSKTLAPKPGPITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 1143  
QY 241 RRGDSRGSLSRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVPIVESHMET 300  
DB 1144 RRGDSRGSLSRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVPIVESHMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 6  
US-09-263-933-4  
; Sequence 4, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/263,933  
; CURRENT FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: 09/129,611  
; EARLIER FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
US-09-263-933-4

Query Match 96.1%; Score 1527; DB 3; Length 1692;

Best Local Similarity 94.4%; Pred. No. 3.5e-144;  
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
DB 183 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYYVDHLTPLODMAHAG 242  
QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 120  
DB 243 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
QY 121 RLAPITAYSOOTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 180  
DB 303 RLAPITAYSOOTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 362  
QY 181 GAGSKTLAPKPGPITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 240  
DB 363 GAGSKTLAPKPGPITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 422  
QY 241 RRGDSRGSLSRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVPIVESHMET 300  
DB 423 RRGDSRGSLSRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKADVPIVESHMET 482  
QY 301 TMR 303  
DB 483 TMR 485

RESULT 7  
US-09-919-901-4  
; Sequence 4, Application US/09919901  
; Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263,933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION:  
US-09-919-901-4

Query Match 96.1%; Score 1527; DB 4; Length 1692;  
Best Local Similarity 94.4%; Pred. No. 3.5e-144;  
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
DB 183 AGITKVPYFVRAOGLIRACMLVRKAAGHYVOMAFMKLAALTGTYYVDHLTPLODMAHAG 242  
QY 61 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 120  
DB 243 LRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREIILGPADNFEQGW 302  
QY 121 RLAPITAYSOOTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 180  
DB 303 RLAPITAYSOOTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVYH 362  
QY 181 GAGSKTLAPKPGPITOMYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTTRHADVIPVR 240

Db 363 GAGSKTLAGEKPIITQWYTNVDOLVGMQAPPGARSILPTCTGSSDLVYTRHADVIPIVR 422  
QY 241 RRGDSRGLSPRPVSYLKSSGGPILCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
Db 423 RRGDSRGLSPRPVSYLKSSGGPILCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 482  
QY 301 TMR 303  
Db 483 TMR 485

RESULT 8  
US-09-263-933-2  
Sequence 2, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT FILING DATE: 1999-03-06  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-2

Query Match 96.1%; Score 1527; DB 3; Length 2307;  
Best Local Similarity 94.4%; Pred. No. 5,4e-144;  
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDMAHAG 60  
Db 275 AGITRVYFVAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDMAHAG 334  
QY 61 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGIPVARRGREIILGPADNPEGQM 120  
Db 335 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGIPVARRGREIILGPADNPEGQM 394  
QY 121 RLAPITAYSQOTRGILGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVYH 180  
Db 395 RLAPITAYSQOTRGILGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVYH 454  
QY 181 GAGSKTLAGEKPIITQWYTNVDOLVGMQAPPGARSILPTCTGSSDLVYTRHADVIPIVR 240  
Db 455 GAGSKTLAGEKPIITQWYTNVDOLVGMQAPPGARSILPTCTGSSDLVYTRHADVIPIVR 514  
QY 241 RRGDSRGLSPRPVSYLKSSGGPILCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
Db 515 RRGDSRGLSPRPVSYLKSSGGPILCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 574  
QY 301 TMR 303  
Db 575 TMR 577

RESULT 9  
US-09-919-901-2  
Sequence 2, Application US/09919901  
Patent No. 6599738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT FILING DATE: 1999-03-06  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-919-901-2

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-2

Query Match 96.1%; Score 1527; DB 4; Length 2307;  
Best Local Similarity 94.4%; Pred. No. 5,4e-144;  
Matches 286; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDMAHAG 60  
Db 275 AGITRVYFVAQGLIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHITPLQDMAHAG 334  
QY 61 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGIPVARRGREIILGPADNPEGQM 120  
Db 335 LRDLAFAVEPVFSDMEVKIITWGADTAACGDIISGIPVARRGREIILGPADNPEGQM 394  
QY 121 RLAPITAYSQOTRGILGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVYH 180  
Db 395 RLAPITAYSQOTRGILGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCTVYH 454  
QY 181 GAGSKTLAGEKPIITQWYTNVDOLVGMQAPPGARSILPTCTGSSDLVYTRHADVIPIVR 240  
Db 455 GAGSKTLAGEKPIITQWYTNVDOLVGMQAPPGARSILPTCTGSSDLVYTRHADVIPIVR 514  
QY 241 RRGDSRGLSPRPVSYLKSSGGPILCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 300  
Db 515 RRGDSRGLSPRPVSYLKSSGGPILCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMET 574  
QY 301 TMR 303  
Db 575 TMR 577

RESULT 10  
US-09-263-933-11  
Sequence 11, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT FILING DATE: 1999-03-06  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-11

Query Match 95.9%; Score 1524; DB 3; Length 1692;  
Best Local Similarity 94.1%; Pred. No. 6,9e-144;  
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

Fri May 7 13:36:59 2004

us-10-650-585-10.rai

Page 5

QY	1	AGTTKPYEFRAAGGLIRACMLYKKAAGGYVMAEMTKLAALGTGVVDHLPLQDANHAG	60
Db	183	AGTRPYEYFRAAGGLIHACMLYKKAAGGYVMAEMTKLALGTGYIYNHLLPLQDANHAG	242
QY	61	LRDLAAVAEVEVITSDMEVKIITWGADTLAACGIIIGLGFVSAARGREILLGPDNFEGQW	120
Db	243	LRDLAAVAEVEVSDMEIKIITWGADTLAACGIIIGLGFVSAARGREILLGPDNSLEBGM	302
QY	121	RLLAFTAYASQOTRGLLGCIIITSLTGKDNQVGESEGVQVSTATQSLATVCVNGCMTVFH	180
Db	303	RLLAFTAYASQOTRGLLGCIIITSLTGKDNQVGESEGVQVSTATQSLATVCVNGCMTVYH	362
QY	181	GAGSKTLTAGKGIITQYMTNVDDLVGMQAPRGASMTPTCGSSDLYLTVIRADVIPIVR	240
Db	363	GAGSKTLTAGKGIITQYMTNVDDLVGMQAPRGASLTPCTGSSDLYLTVIRADVIPIVR	422
QY	241	RQDPSRGSLLSPRPVSYLKGSGGPIILCPSHAVGIFPAAVCTRGVAKXAVDPIVESMET	300
Db	423	RQDPSRGSLLSPRPVSYLKGSAAGGPILCPSHAVGIFPAAVCTRGVAKXAVDPIVESMET	482
QY	301	TMR	
Db	483	TMR 485	

```

RESULT 11
US-09-919-901-11
: Sequence 11, Application US/09919901
: Patent No. 6599738
:
GENERAL INFORMATION:
:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patick, Amy K.
: TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
: TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
:
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
:
PRIOR APPLICATION NUMBER: 09/363,933
PRIOR FILING DATE: 1999-02-08
:
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
:
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
:
SEQ ID NO 11
:
LENGTH: 1692
:
TYPE: PRT
:
ORGANISM: Artificial Sequence
FEATURE:
:
OTHER INFORMATION: :
US-09-919-901-11

```

[illegible]

Db 423 RRDSNGSLSPRPVSYLKGSA G3LLCPGHA VGFPAACVCTRGVAKAVDFVVESEMET 482  
QY 301 TMR 303  
Db 483 TMR 485

```

RESULT 12
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patlick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

```

```

Query Match: 95.9%; Score 1524; DB 3; Length 2307;
Best Local Similarity: 94.1%; Pred. No. 1, 1e-143;
Matches: 285; Conservative: 12; Mismatches: 6; Indels: 0; Gaps: 0;

QY      1 AGITTYEYFVRAQGLIRACMLVKKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db      275 AGITRVYFVRAQGLIRACMLVKKVAGGHVQMAFMKLGALTGTYYVYNHLLTPLRDMAHAG 334

QY      61 LRLAVAVEVITSDMEVKIITWGADPAAAGDIIISGLPVARRGRELLGPADNFEQGW 120
Db      335 LRLDAVAEVEVVSDEETKIITWGADPAAAGDIIISGLPVARRKETLLGPADSLBEGGW 394

QY      121 RLAPITAYSQQRGLGLGCIITSLTGRDKQVGEVGVVSTATQSFATCNGVCWTFVH 180
Db      395 RLAPITAYSQQRGLGLGCIITSLTGRDKQVGEVGVVSTATQSFATCNGVCWTFVH 454

QY      181 GAGSKTLAGEKPIITQMTYTNVDODLVGMQAPPGARSMTPCTCGSSDLIYVTRHADVIPVR 240
Db      455 GAGSKTLAGEKPIITQMTYTNVDODLVGMQAPPGARSLTPCTCGSSDLIYVTRHADVIPVR 514

QY      241 RRGDSRSLSPRPVSLKGSAGSGLLCPSGHVGIFRAAVCTRGVAKAVDFVESMET 300
Db      515 RRGDSRSLSPRPVSLKGSAGSGLLCPSGHVGIFRAAVCTRGVAKAVDFVESMET 574

QY      301 TMR 303
Db      575 TMR 577

RESULT 13
US-09-919-901-9
: Sequence 9, Application US/09919901
: Patent No. 6599738
: GENERAL INFORMATION:
: APPLICANT: Potts, Karen E.
: APPLICANT: Jackson, Roberta L.
: TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
: TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
: FILE REFERENCE: 0125-0005A
: CURRENT APPLICATION NUMBER: US/09/919,901
: CURRENT PRIORITY NUMBER: 002-0002

```

PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-9

Query Match 95.3%; Score 1524; DB 4; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 1,1e-143;  
Matches 285; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLODWAHAG 60  
DB 275 AGITRVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLGALTGTYYVNHLLTPLRDWAHAG 334  
QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRRIILGPADNFEQGW 120  
DB 335 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGKELILGPADNFEGRGW 394  
QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVYH 180  
DB 395 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVYH 454  
QY 181 GAGSKTLAAGPKGPIITQWYTNVDODLVGMOAPPGARSMTPTCTGSSDLVYVTRHADVTPVR 240  
DB 455 GAGSKTLAAGPKGPIITQWYTNVDODLVGMOAPPGARSMTPTCTGSSDLVYVTRHADVTPVR 514  
QY 241 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFVVEESMET 300  
DB 515 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFVVEESMET 574  
QY 301 TMR 303  
DB 575 TMR 577

RESULT 14  
US-09-539-601-33  
Sequence 33, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Barteneschlager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
EARLIER FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 33  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-33

Query Match 95.8%; Score 1523; DB 4; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 2e-143;  
Matches 287; Conservative 8; Mismatches 8; Indels 0; Gaps 0;  
QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLODWAHAG 60  
DB 904 AGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGTYVDHLTPLRDWAHAG 963  
QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRRIILGPADNFEQGW 120

DB 964 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRRIILGPADNFEQGW 1023  
QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVYH 180  
DB 1024 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVYH 1083  
QY 181 GAGSKTLAAGPKGPIITQWYTNVDODLVGMOAPPGARSMTPTCTGSSDLVYVTRHADVTPVR 240  
DB 1084 GAGSKTLAAGPKGPIITQWYTNVDODLVGMOAPPGARSMTPTCTGSSDLVYVTRHADVTPVR 1143  
QY 241 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFVVEESMET 300  
DB 1144 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFVVEESMET 1203  
QY 301 TMR 303  
DB 1204 TMR 1206

RESULT 15  
US-09-263-933-18  
Sequence 18, Application US/09263933  
Patent No. 6280340  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
EARLIER FILING DATE: 1999-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 18  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-18

Query Match 95.3%; Score 1515; DB 3; Length 1692;  
Best Local Similarity 93.7%; Pred. No. 5,6e-143;  
Matches 284; Conservative 12; Mismatches 7; Indels 0; Gaps 0;  
QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYVDHLTPLODWAHAG 60  
DB 183 AGITRVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLGALTGTYYVNHLLTPLRDWAHAG 242  
QY 61 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGRRIILGPADNFEQGW 120  
DB 243 LRDLAAVEPVFSDMEVKIITWGADTAACGDIISGLPVSARRGKELILGPADNFEGRGW 302  
QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVYH 180  
DB 303 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSPLATCVNGVCTVYH 362  
QY 181 GAGSKTLAAGPKGPIITQWYTNVDODLVGMOAPPGARSMTPTCTGSSDLVYVTRHADVTPVR 240  
DB 363 GAGSKTLAAGPKGPIITQWYTNVDODLVGMOAPPGARSMTPTCTGSSDLVYVTRHADVTPVR 422  
QY 241 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFVVEESMET 300  
DB 423 RRGDSRGSLLSPRVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFVVEESMET 482  
QY 301 TMR 303  
DB 483 TMR 485

Search completed: May 6, 2004, 09:39:01  
Job time: 12.8923 secs



2 Fri May 7 13:36:59 2004

us-10-650-585-10.rai

---

Page 7



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 : Search time 30.9958 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-10

Sequence: 1 AGITKVPYVRAQGLIRACM.....RGVAKAVDFIPVSMETTR 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCR\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCRUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1589	100.0	303	US-10-017-736-10	Sequence 10, Appl
2	1589	100.0	303	US-10-650-585-10	Sequence 10, Appl
3	1589	100.0	334	US-10-017-736-4	Sequence 4, Appl1
4	1589	100.0	334	US-10-650-585-4	Sequence 4, Appl1
5	1589	100.0	341	US-10-017-736-14	Sequence 14, Appl
6	1589	100.0	341	US-10-650-585-14	Sequence 14, Appl
7	1589	100.0	352	US-10-017-736-13	Sequence 13, Appl
8	1589	100.0	352	US-10-650-585-13	Sequence 13, Appl
9	1589	100.0	380	US-10-017-736-12	Sequence 12, Appl
10	1589	100.0	380	US-10-650-585-12	Sequence 12, Appl
11	1589	100.0	393	US-10-017-736-11	Sequence 11, Appl
12	1589	100.0	393	US-10-650-585-11	Sequence 11, Appl
13	1589	100.0	409	US-10-017-736-2	Sequence 2, Appl1
14	1589	100.0	409	US-10-650-585-2	Sequence 2, Appl1
15	1580	99.4	303	US-10-017-736-18	Sequence 18, Appl

16	1580	99.4	303	US-10-650-585-18	Sequence 18, Appl
17	1579	99.4	303	US-10-017-736-16	Sequence 16, Appl
18	1579	99.4	303	US-10-650-585-16	Sequence 16, Appl
19	1570	98.8	301	US-10-017-736-17	Sequence 17, Appl
20	1570	98.8	301	US-10-650-585-17	Sequence 17, Appl
21	1532	96.4	292	US-10-017-736-15	Sequence 15, Appl
22	1532	96.4	292	US-10-650-585-15	Sequence 15, Appl
23	1530	96.3	2201	US-10-029-907-3	Sequence 3, Appl1
24	1530	96.3	2201	US-10-029-907-3	Sequence 3, Appl1
25	1530	96.3	3010	US-10-467-000-1	Sequence 1, Appl
26	1527	96.1	1692	US-10-191-866-4	Sequence 4, Appl1
27	1527	96.1	1692	US-10-191-866-4	Sequence 4, Appl1
28	1527	96.1	2307	US-09-919-901-2	Sequence 2, Appl1
29	1527	96.1	2307	US-10-191-866-2	Sequence 2, Appl1
30	1524	95.9	1692	US-09-919-901-11	Sequence 11, Appl
31	1524	95.9	1692	US-10-191-866-11	Sequence 11, Appl
32	1524	95.9	2307	US-09-919-901-9	Sequence 9, Appl1
33	1524	95.9	2307	US-10-191-866-9	Sequence 9, Appl1
34	1515	95.3	1692	US-09-919-901-18	Sequence 18, Appl
35	1515	95.3	1692	US-10-191-866-18	Sequence 18, Appl
36	1515	95.3	2307	US-09-919-901-16	Sequence 16, Appl
37	1515	95.3	2307	US-10-191-866-16	Sequence 16, Appl
38	1478	93.0	2201	US-10-085-476-2	Sequence 2, Appl1
39	1405	88.4	3011	US-09-742-659-4	Sequence 4, Appl1
40	1405	88.4	3011	US-09-881-894-3	Sequence 3, Appl1
41	1405	88.4	3011	US-10-184-150-3	Sequence 3, Appl1
42	1405	88.4	3011	US-10-328-997-3	Sequence 3, Appl1
43	1405	88.4	3012	US-09-238-076-2	Sequence 2, Appl1
44	1405	88.4	3012	US-09-995-937-2	Sequence 2, Appl1
45	1405	88.4	3012	US-09-917-563-2	Sequence 2, Appl1

#### ALIGNMENTS

RESULT 1  
US-10-017-736-10  
; Sequence 10, Application US/10017736  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 10  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-10

Query Match 100.0%; Score 1589; DB 13; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1.5e+154;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AGITKVPYVRAQGLIRACM	LVKKAAGHYVMAEMKLAALTGTYYVHTLPLQMAHNG	60
DB	1	AGITKVPYVRAQGLIRACM	LVKKAAGHYVMAEMKLAALTGTYYVHTLPLQMAHNG	60
QY	61	LRDLAAVPEVPSDMEVKIITWGADTAACGDIISGLPVSA	RGRSEIILGPADNFEQCGM	120
DB	61	LRDLAAVPEVPSDMEVKIITWGADTAACGDIISGLPVSA	RGRSEIILGPADNFEQCGM	120
QY	121	RLIAPITAVSQOTRGLIGCIITSLGRDNQVEGEVQVSTATQSP	FLATCVNGVCTVTFH	180
DB	121	RLIAPITAVSQOTRGLIGCIITSLGRDNQVEGEVQVSTATQSP	FLATCVNGVCTVTFH	180
QY	181	GAGSKTLA	PKXPIITQMTNNVDQDLYGWAQPARGASMPCTCGSSDLVYVTHADIVPR	240

Db 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 240  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 QY 301 TMR 303  
 Db 301 TMR 303

RESULT 2  
 US-10-650-585-10  
 ; Sequence 10, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 10  
 ; LENGTH: 303  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-10

Query Match 100.0%; Score 1589; DB 16; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 QY 61 LRDLAVALVEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 61 LRDLAVALVEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
 Db 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
 QY 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 240  
 Db 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 240  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 QY 301 TMR 303  
 Db 301 TMR 303

RESULT 3  
 US-10-017-736-4  
 ; Sequence 4, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 334  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-017-736-4

Query Match 100.0%; Score 1589; DB 13; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 16 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 QY 61 LRDLAVALVEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 76 LRDLAVALVEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 135  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180  
 Db 136 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 195  
 QY 181 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 240  
 Db 196 GAGSKTLAAGPKGPIITQMTYNTVDOLVGMQAPPGARSMTPCTCGSSDLVLTTRHADVIPIVR 255  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 256 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 315  
 QY 301 TMR 303  
 Db 316 TMR 318

RESULT 4  
 US-10-650-585-4  
 ; Sequence 4, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 334  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-4

Query Match 100.0%; Score 1589; DB 16; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 16 AGITKVPYFVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 75  
 QY 61 LRDLAVALVEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 76 LRDLAVALVEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 135  
 QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVPH 180

```

Db 136 RLAPITAVSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCVNGVCTVPH 195
QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMTPTCCSSDLVYTRHADVIPIVR 240
Db 196 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMTPTCCSSDLVYTRHADVIPIVR 255
QY 241 RRGDSRGSLLSPRVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 300
Db 256 RRGDSRGSLLSPRVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 315
QY 301 TMR 303
Db 316 TMR 318

```

```

RESULT 5
US-10-017-736-14
; Sequence 14, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-14

```

```

Query Match 100.0%; Score 1589; DB 13; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.7e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db 39 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
QY 61 LRDLAAVAVPEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 120
Db 99 LRDLAAVAVPEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 158
QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCVNGVCTVPH 180
Db 159 RLAPITAVSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCVNGVCTVPH 218
QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMTPTCCSSDLVYTRHADVIPIVR 240
Db 219 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMTPTCCSSDLVYTRHADVIPIVR 278
QY 241 RRGDSRGSLLSPRVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 300
Db 279 RRGDSRGSLLSPRVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 358
QY 301 TMR 303
Db 339 TMR 341

```

```

RESULT 6
US-10-650-585-14
; Sequence 14, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082

```

```

; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-14

```

```

Query Match 100.0%; Score 1589; DB 16; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.7e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db 39 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 98
QY 61 LRDLAAVAVPEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 120
Db 99 LRDLAAVAVPEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 158
QY 121 RLAPITAVSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCVNGVCTVPH 180
Db 159 RLAPITAVSQOTRGLGCIITSLTGRDKQVEGEVQVSTATQSFPLATCVNGVCTVPH 218
QY 181 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMTPTCCSSDLVYTRHADVIPIVR 240
Db 219 GAGSKTLAGPKPITQMTYTNVDQDLVGWQAPPGARSMTPTCCSSDLVYTRHADVIPIVR 278
QY 241 RRGDSRGSLLSPRVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 300
Db 279 RRGDSRGSLLSPRVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMET 338
QY 301 TMR 303
Db 339 TMR 341

```

```

RESULT 7
US-10-017-736-13
; Sequence 13, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 352
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-13

```

```

Query Match 100.0%; Score 1589; DB 13; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.8e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60
Db 50 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 109
QY 61 LRDLAAVAVPEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGRIILGPADNFEQGM 120

```

Db 110 LRDIAVAEVEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 169  
 QY 121 RLAPITAYVSOQTRGLIGCIITSLTGRDKNOVEGEVQVNSTAQSLATCNGVCMTVFH 180  
 Db 170 RLAPITAYVSOQTRGLIGCIITSLTGRDKNOVEGEVQVNSTAQSLATCNGVCMTVFH 229  
 QY 181 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIR 240  
 Db 230 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIR 289  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGCPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 290 RRGDSRGSLLSPRPVSYLKSSGCPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 349  
 QY 301 TMR 303  
 Db 350 TMR 352

## RESULT 8

US-10-650-585-13  
 ; Sequence 13, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-13

Query Match 100.0%; Score 1589; DB 16; Length 352;

Best Local Similarity 100.0%; Pred. No. 1.8e-154;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 50 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 109  
 QY 61 LRDIAVAEVEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 110 LRDIAVAEVEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 169  
 QY 121 RLAPITAYVSOQTRGLIGCIITSLTGRDKNOVEGEVQVNSTAQSLATCNGVCMTVFH 180  
 Db 170 RLAPITAYVSOQTRGLIGCIITSLTGRDKNOVEGEVQVNSTAQSLATCNGVCMTVFH 229  
 QY 181 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIR 240  
 Db 230 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIR 289  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGCPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 290 RRGDSRGSLLSPRPVSYLKSSGCPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 349  
 QY 301 TMR 303  
 Db 350 TMR 352

## RESULT 9

US-10-017-736-12  
 ; Sequence 12, Application US/10017736

; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-017-736-12

Query Match 100.0%; Score 1589; DB 13; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 2e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 Db 78 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 137  
 QY 61 LRDIAVAEVEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 120  
 Db 138 LRDIAVAEVEVIFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGM 197  
 QY 121 RLAPITAYVSOQTRGLIGCIITSLTGRDKNOVEGEVQVNSTAQSLATCNGVCMTVFH 180  
 Db 198 RLAPITAYVSOQTRGLIGCIITSLTGRDKNOVEGEVQVNSTAQSLATCNGVCMTVFH 257  
 QY 181 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIR 240  
 Db 258 GAGSKTLAGPKGPIITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLVLTTRADVIPIR 317  
 QY 241 RRGDSRGSLLSPRPVSYLKSSGCPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 Db 318 RRGDSRGSLLSPRPVSYLKSSGCPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 377  
 QY 301 TMR 303  
 Db 378 TMR 380

## RESULT 10

US-10-650-585-12  
 ; Sequence 12, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-12

Query Match 100.0%; Score 1589; DB 16; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 2e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60

```

Db      ||||| 78 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQNAFMKLAALTGYVDHLTPLODMAHAG 137
Qy      ||||| 61 LRDLAVALVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFEQGM 120
Db      ||||| 138 LRDLAVALVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFEQGM 197
Qy      ||||| 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVFFH 180
Db      ||||| 198 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVFFH 257
Qy      ||||| 181 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 240
Db      ||||| 258 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 317
Qy      ||||| 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
Db      ||||| 318 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 377
Qy      ||||| 301 TMR 303
Db      ||||| 378 TMR 380

```

# RESULT 11

```

; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

```

```

Query Match      100.0%; Score 1589; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.1e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      ||||| 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQNAFMKLAALTGYVDHLTPLODMAHAG 60
Db      ||||| 91 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQNAFMKLAALTGYVDHLTPLODMAHAG 150
Qy      ||||| 61 LRDLAVALVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFEQGM 120
Db      ||||| 151 LRDLAVALVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFEQGM 210
Qy      ||||| 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVFFH 180
Db      ||||| 211 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVFFH 270
Qy      ||||| 181 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 240
Db      ||||| 271 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 330
Qy      ||||| 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
Db      ||||| 331 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 390
Qy      ||||| 301 TMR 303
Db      ||||| 391 TMR 393

```

# RESULT 12

```

US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR FILING DATE: US/10/017,736A
; PRIOR APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

```

```

Query Match      100.0%; Score 1589; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.1e-154;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      ||||| 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQNAFMKLAALTGYVDHLTPLODMAHAG 60
Db      ||||| 91 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQNAFMKLAALTGYVDHLTPLODMAHAG 150
Qy      ||||| 61 LRDLAVALVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFEQGM 120
Db      ||||| 151 LRDLAVALVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFEQGM 210
Qy      ||||| 121 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVFFH 180
Db      ||||| 211 RLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFILATCVNGVCTVFFH 270
Qy      ||||| 181 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 240
Db      ||||| 271 GAGSKTLAAGPKGPIITOMYTNVDQDLVGMQAPPGARSMTPTCTGSSDLVYTRHADVIPIVR 330
Qy      ||||| 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 300
Db      ||||| 331 RRGDSRGSLLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMET 390
Qy      ||||| 301 TMR 303
Db      ||||| 391 TMR 393

```

# RESULT 13

```

US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2

```

```

Query Match      100.0%; Score 1589; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 2.3e-154;

```

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 DB 95 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 154

QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRRREIILGPADNFEQCGW 120  
 DB 155 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRRREIILGPADNFEQCGW 214

QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCTVPH 180  
 DB 215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCTVPH 274

QY 181 GAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADYIPVR 240  
 DB 275 GAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADYIPVR 334

QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 335 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

QY 301 TMR 303  
 DB 395 TMR 397

RESULT 14  
 US-10-650-585-2  
 ; Sequence 2, Application US/10650585  
 ; Publication No. US2004007066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 409  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-650-585-2

Query Match 100.0%; Score 1589; DB 16; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 2,3e-154;  
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 DB 95 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 154

QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRRREIILGPADNFEQCGW 120  
 DB 155 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRRREIILGPADNFEQCGW 214

QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCTVPH 180  
 DB 215 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCTVPH 274

QY 181 GAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADYIPVR 240  
 DB 275 GAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADYIPVR 334

QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 335 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 394

QY 301 TMR 303  
 DB 395 TMR 397

RESULT 15  
 US-10-017-736-18  
 ; Sequence 18, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 18  
 ; LENGTH: 303  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 US-10-017-736-18

Query Match 99.4%; Score 1580; DB 13; Length 303;  
 Best Local Similarity 99.7%; Pred. No. 1.2e-153;  
 Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60  
 DB 1 AGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAG 60

QY 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRRREIILGPADNFEQCGW 120  
 DB 61 LRLDAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVARSRRREIILGPADNFEQCGW 120

QY 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCTVPH 180  
 DB 121 RLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCNGVCTVPH 180

QY 181 GAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADYIPVR 240  
 DB 181 GAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLYLTRHADYIPVR 240

QY 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300  
 DB 241 RRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMET 300

QY 301 TMR 303  
 DB 301 TMR 303

Search completed: May 6, 2004, 09:43:18  
 Job time : 30.9958 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 53.4939 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MAASCGAVFVIGLALTLSP.....RGVAKAVDFVPSMETMR 393

Scoring table: BIOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_290Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	393	5	ABG32184 HCV prote
2	2053	100.0	409	5	ABG32181 HCV prote
3	1987	96.8	380	5	ABG32185 HCV prote
4	1967	95.8	3010	2	AAR82694 Partial H
5	1962	95.6	3010	2	AAR82622 HCV prote
6	1961	95.5	3010	2	AAR68864 Hepatitis
7	1951	95.0	2201	5	ABG30591 Hepatitis
8	1951	95.0	2201	5	ABG30591 Hepatitis
9	1951	95.0	2201	5	ABG30600 Hepatitis
10	1951	95.0	2201	5	ABG30581 Hepatitis
11	1951	95.0	2201	5	ABG30593 Hepatitis
12	1951	95.0	2201	5	ABG30582 Hepatitis
13	1951	95.0	2201	5	ABG30580 Hepatitis
14	1951	95.0	2201	5	ABG30587 Hepatitis
15	1951	95.0	2201	5	ABG30599 Hepatitis
16	1951	95.0	2201	5	ABG30594 Hepatitis
17	1951	95.0	2201	5	ABG30598 Hepatitis
18	1951	95.0	2201	5	ABG30595 Hepatitis
19	1951	95.0	3010	5	ABG32458 Hepatitis
20	1951	95.0	3010	5	ABG32459 Hepatitis
21	1951	95.0	3010	5	ABG32451 Hepatitis
22	1951	95.0	3010	5	ABG32455 Hepatitis
23	1951	95.0	3010	5	ABG32457 Hepatitis
24	1951	95.0	3010	5	ABG32460 Hepatitis
25	1951	95.0	3010	5	ABG32461 Hepatitis

26	1951	95.0	3010	5	ABG32454 Hepatitis
27	1951	95.0	3011	5	ABG32456 Hepatitis
28	1948	94.9	2201	5	ABG30586 Hepatitis
29	1948	94.9	2201	5	ABG30589 Hepatitis
30	1948	94.9	2201	5	ABG30583 Hepatitis
31	1948	94.9	2201	5	ABG30588 Hepatitis
32	1947	94.8	2201	5	ABG30590 Hepatitis
33	1946	94.8	2307	3	AAY70064 Recombina
34	1945	94.7	3010	5	ABG32452 Hepatitis
35	1944	94.7	2201	5	ABG30584 Hepatitis
36	1944	94.7	2201	5	ABG30602 Hepatitis
37	1944	94.7	3010	5	ABG32453 Hepatitis
38	1943	94.6	2307	3	AAY70065 Recombina
39	1943	94.6	3014	2	AAR54099 NANBHV E1
40	1940	94.5	2201	5	ABG30585 Hepatitis
41	1938.5	94.4	768	2	AAR40223 Recombina
42	1938	94.4	3014	2	AAR35207 Recombina
43	1937	94.3	3090	7	ADD67962 EMCV infe
44	1934	94.2	2307	3	AAY70066 Recombina
45	1929	94.0	3010	5	AAR20477 HCV-S1 fu

## ALIGNMENTS

RESULT 1  
ID ABG32184 standard; protein; 393 AA.  
XX  
AC ABG32184;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 815-1206.  
XX  
KM HCV, enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutain.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200248375-A2.  
XX  
PD 20-JUN-2002.  
XX  
PF 13-DEC-2001; 2001WO-CA001796.  
XX  
PR 15-DEC-2000; 2000US-0256031P.  
XX  
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX  
PS Claim 41; Page 59-60; 67pp; English.  
XX  
CC The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX SQ Sequence 393 AA;

Query Match 100.0%; Score 2053; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 7, 1e-191; Indels 0; Gaps 0;  
 Matches 393; Conservative 0; Mismatches 0;

QY 1 MAASCGAVFTGLALTLSPYKYLARLIMWLQYLITRFAHLQWIPPLNVGGRDAI 60  
 DB 1 MAASCGAVFTGLALTLSPYKYLARLIMWLQYLITRFAHLQWIPPLNVGGRDAI 60  
 QY 61 ILITCAHPELIPDITLLAIFGLPMLVQAGITKPYEFAAGILRACMLYKAGAHY 120  
 DB 61 ILITCAHPELIPDITLLAIFGLPMLVQAGITKPYEFAAGILRACMLYKAGAHY 120  
 QY 121 VQMAFMALALGTYYVDHLPLQDMAHAGLRDLAVAVEVIPSDEVKITIGADTAA 180  
 DB 121 VQMAFMALALGTYYVDHLPLQDMAHAGLRDLAVAVEVIPSDEVKITIGADTAA 180  
 QY 181 GDIISGLPVAGARRELLGPADNFBQGRLLAPTAAYSQQRGLGCIITSLTRDKN 240  
 DB 181 GDIISGLPVAGARRELLGPADNFBQGRLLAPTAAYSQQRGLGCIITSLTRDKN 240  
 QY 181 GDIISGLPVAGARRELLGPADNFBQGRLLAPTAAYSQQRGLGCIITSLTRDKN 240  
 DB 181 GDIISGLPVAGARRELLGPADNFBQGRLLAPTAAYSQQRGLGCIITSLTRDKN 240  
 QY 241 QVEGEVQVSTATQSFATCVNGCWTVEHAGSKTLAGKGPITQMTYVNDODLVGWA 300  
 DB 241 QVEGEVQVSTATQSFATCVNGCWTVEHAGSKTLAGKGPITQMTYVNDODLVGWA 300  
 QY 301 PPGARSTPCTCGSSDLYLTVTRADVTPVRRGDSGSLSPVATLKSSGGPLCP 360  
 DB 301 PPGARSTPCTCGSSDLYLTVTRADVTPVRRGDSGSLSPVATLKSSGGPLCP 360  
 QY 361 GHAVGIFRAAVCTRGAVKAVDFIPVESMETTMR 393  
 DB 361 GHAVGIFRAAVCTRGAVKAVDFIPVESMETTMR 393

RESULT 2  
 ABG32181  
 ID ABG32181 standard; protein; 409 AA.

XX AC ABG32181;  
 XX DT 05-NOV-2002 (first entry)  
 XX DE HCV protease NS2/3 (810-1206).  
 XX HCV protease NS2/3 (810-1206); hepatitis C virus infection;  
 KW chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KW hepatocellular carcinoma; hepatitis; hepatitis B virus infection;  
 KW chaotropic agent; mutant; mutagen.

XX OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 398..409  
 FT /note="Streptavidin tag"  
 PN WO200248375-A2.  
 PD 20-UN-2002.  
 PF 13-DEC-2001; 2001WO-CA001796.  
 PR 15-DEC-2000; 2000US-0256031P.  
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI: 2002-599511/64.  
 DR N-PSDB; ABR90406.  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

Claim 42; Fig 1B; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 residue amino acid 810 to 906, or having a minimal amino acid sequence  
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 NS2/3 protease. Also included are (1) a composition (C) comprising an  
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 its truncation or a mutated sequence, where the protease is in a solution  
 comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 protease, involving isolating the protease in the presence of a  
 chaotropic agent, refolding the isolated protease by contacting it with a  
 reducing agent, and LDAO in the presence of reduced concentration of the  
 chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 protease, involving diluting refolded inactive NS2/3 protease in a medium  
 containing an activation detergent to induce auto-cleavage of the NS2/3  
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 protease, involving incubating the active NS2/3 protease produced by M2  
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 cleavage products or their fragments, and measuring the presence or  
 absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 active NS2/3 protease, involving carrying out M3 in the presence of, or  
 absence of the potential inhibitor, comparing the amount of uncleaved  
 NS2/3 protease, cleavage products or their fragments. The protease is  
 useful for detailed biochemical characterisation of the enzymes and in  
 the development of in vitro assays for screening novel inhibitors of  
 NS2/3 protease which are useful as therapeutic agents against HCV  
 infection (which causes chronic liver disease, cirrhosis and end-stage  
 liver disease. M1 is useful for high level production of protease. The  
 present sequence represents the NS2/3 (810-1206) protein, which has a C-  
 terminal streptavidin tag

XX SQ Sequence 409 AA;

Query Match 100.0%; Score 2053; DB 5; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 7, 1e-191; Indels 0; Gaps 0;  
 Matches 393; Conservative 0; Mismatches 0;

QY 1 MAASCGAVFTGLALTLSPYKYLARLIMWLQYLITRFAHLQWIPPLNVGGRDAI 60  
 DB 5 MAASCGAVFTGLALTLSPYKYLARLIMWLQYLITRFAHLQWIPPLNVGGRDAI 64

QY 61 ILITCAVHPELIDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVKAAGHY 120  
 DB 65 ILITCAVHPELIDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVKAAGHY 124  
 QY 121 VQMAFMKLAALITGVYDHLTPQDMAHGLRLAVALVEVIFSDMEVKITWGAUTAAC 180  
 DB 125 VQMAFMKLAALITGVYDHLTPQDMAHGLRLAVALVEVIFSDMEVKITWGAUTAAC 184  
 QY 181 GDIIISGIPVARSRRGELLGPADNFEQGGWRLLAPITAVSQOTRGILGCIITSLTRDKN 240  
 DB 185 GDIIISGIPVARSRRGELLGPADNFEQGGWRLLAPITAVSQOTRGILGCIITSLTRDKN 244  
 QY 241 QVGEVQVYSTAQSFLATCVNGVCTVFHGAASKTLAGKGIITQMTNVDDLVGMOA 300  
 DB 245 QVGEVQVYSTAQSFLATCVNGVCTVFHGAASKTLAGKGIITQMTNVDDLVGMOA 304  
 QY 301 PFGARSMTPCTCGSSDLVLTVRADVTPVRRGDSRGLSPRPVSYLKGSSGGPILCS 360  
 DB 305 PFGARSMTPCTCGSSDLVLTVRADVTPVRRGDSRGLSPRPVSYLKGSSGGPILCS 364  
 QY 361 GHAIVGFRAAVCTRGVAKAVDFIPVBSMETTMR 393  
 DB 365 GHAIVGFRAAVCTRGVAKAVDFIPVBSMETTMR 397

RESULT 3  
 ABG32185  
 ID ABG32185 standard; protein; 380 AA.  
 XX  
 AC ABG32185;

XX 05-NOV-2002 (first entry)  
 DE HCV protease NS2/3 truncation mutant 827-1206.  
 KM HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethyamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutagen.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 PN WO200248375-A2.  
 PD 20-JUN-2002.  
 XX 13-DEC-2001; 2001WC-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 PA Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-539511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 60-61; 67pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)) or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethyamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease, (2) a NS2/3 inhibitory peptide

CC appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation inhibitor to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)

SO Sequence 380 AA;

Query Match 96.8%; Score 1987; DB 5; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-184;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ALTLSPYKYLALRIWMLOYLITRVAHLOVPIPNVAGGRPAITLTCVHPELIF 73  
 DB 1 ALTLSPYKYLALRIWMLOYLITRVAHLOVPIPNVAGGRPAITLTCVHPELIF 60  
 QY 74 DITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 133  
 DB 6 DITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 120  
 QY 134 TTYVDHLTPQDMAHGLRLAVALVEVIFSDMEVKITWGAUTAACDIIISGLVSR 193  
 DB 121 TTYVDHLTPQDMAHGLRLAVALVEVIFSDMEVKITWGAUTAACDIIISGLVSR 180  
 QY 194 GREIILGPADNFEQGGWRLLAPITAVSQOTRGILGCIITSLTRDKNVEGEVYSTAT 253  
 DB 181 GREIILGPADNFEQGGWRLLAPITAVSQOTRGILGCIITSLTRDKNVEGEVYSTAT 240  
 QY 254 QGFLLATCVNGVCTVFHGAASKTLAGKGIITQMTNVDDLVGMOAPFGARSMTPCTCG 313  
 DB 241 QGFLLATCVNGVCTVFHGAASKTLAGKGIITQMTNVDDLVGMOAPFGARSMTPCTCG 300  
 QY 314 SSDLVLTVRADVTPVRRGDSRGLSPRPVSYLKGSSGGPILCSGHAIVGFRAAVCT 373  
 DB 301 SSDLVLTVRADVTPVRRGDSRGLSPRPVSYLKGSSGGPILCSGHAIVGFRAAVCT 360  
 QY 374 RGAVARVDFIPVBSMETTMR 393  
 DB 361 RGAVARVDFIPVBSMETTMR 380  
 RESULT 4  
 AAR82694  
 ID AAR82694 standard; protein; 3010 AA.  
 XX  
 AC AAR82694;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 14-NOV-1996 (first entry)  
 XX  
 DE Partial HCV non-structural polypeptide.  
 XX  
 KM proteinase; hepatitis C virus; screening; inhibitor; proteolytic;  
 KM identification; cleavage.  
 XX

OS Hepatitis C virus; Virus.  
 XX Key Location/Qualifiers  
 FH Protein 898..1233  
 FT /note="partial proteinase; see AAR82692"  
 FT Protein 992..1907  
 FT /note="partial proteinase; see AAR82693"  
 XX JP07184648-A.  
 XX  
 XX 25-JUL-1995.  
 XX  
 XX 05-FEB-1993; 93JP-00018854.  
 XX  
 XX 07-FEB-1992; 92JP-00022657.  
 XX 18-SEP-1992; 92JP-00249240.  
 XX 04-DEC-1992; 92JP-00325303.  
 XX  
 XX (KAENNO K.  
 XX (SUMO) SUMITOMO METAL IND LTD.  
 XX (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX  
 XX WPI, 1995-287962/38.  
 XX N-PSDB; AAT03960.  
 XX  
 XX An HCV proteinase active substance - which has activity as an anti-HCV  
 XX agent and can be used to screen for proteinase inhibitors.  
 XX  
 XX Disclosure; Page 39-48; 52pp; Japanese.  
 XX  
 XX The present sequence is a partial Hepatitis C Virus (HCV) polyprotein  
 XX from the non-structural region. Partial proteinase sequences (AAR82692-  
 XX 93) are contained within this sequence. The proteinases can be used as  
 XX anti-HCV agents. They can also be used to screen cpts. for their ability  
 XX to inhibit their proteolytic activity. In this way proteinase inhibitors  
 XX can be identified. (Updated on 16-OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 3010 AA;  
 XX  
 XX Query Match 95.8%; Score 1967; DB 2; Length 3010;  
 XX Best Local Similarity 93.9%; Pred. No. 3,1e-181;  
 XX Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;  
 XX  
 XX 1 MAASCGGAVFTGLALTLSPYKYLRLIMWLQYLITRVEAHQVWVPLNVRGGRDAI 60  
 XX 814 MAASCGGAVFTGLALTLSPYKYLRLIMWLQYLITRVEAHQVWVPLNVRGGRDAI 873  
 XX  
 XX 61 ILTCAVPELIPITIKLLAIFGRLMVLQAGITKVFYFRAQGLIRACMLVRKAGHY 120  
 XX 874 ILTCAVPELIPITIKLLAIFGRLMVLQAGITKVFYFRAQGLIRACMLVRKAGHY 933  
 XX  
 XX 121 VQAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAAVEVPVPSDMEVKIITWADTAAC 180  
 XX 934 VQAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAAVEVPVPSDMEVKIITWADTAAC 993  
 XX  
 XX 181 GDIISGLPVSARRREILGPADNFEQGGWRLAPITAVSQOTRGLGCIITSITGRDN 240  
 XX 994 GDIISGLPVSARRREILGPADNFEQGGWRLAPITAVSQOTRGLGCIITSITGRDN 1053  
 XX  
 XX 241 QVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGKGPITQMTYTNVDQDLVQWA 300  
 XX 1054 QVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGKGPITQMTYTNVDQDLVQWA 1113  
 XX  
 XX 301 PPGARSMTPCTCGSSDLVYLRHADVPVRRRGSRSGLSPRVSYLKSSSGGFLCPS 360  
 XX 1114 PPGARSMTPCTCGSSDLVYLRHADVPVRRRGSRSGLSPRVSYLKSSSGGFLCPS 1173  
 XX  
 XX 361 GHAVGIFRAVCTRGVAKAVDFIVESMETTMR 393  
 XX 1174 GHAVGIFRAVCTRGVAKAVDFIVESMETTMR 1206  
 XX  
 XX RESULT 5

AAR86822  
 ID AAR86822 standard; protein; 3010 AA.  
 XX  
 XX AC AAR86822;  
 XX  
 XX 16-OCT-2003 (revised)  
 XX 16-OCT-1995 (first entry)  
 XX  
 XX HCV protein cleavable with new serine proteinase.  
 XX  
 XX proteinase; serine, cleavage; hepatitis C virus; HCV.  
 XX  
 XX Hepatitis C virus; Virus.  
 XX  
 XX OS  
 XX  
 XX Key Location/Qualifiers  
 XX FT Cleavage-site 2419..2420  
 XX FT /note="Serine protease cleavage site"  
 XX  
 XX JP06315377-A.  
 XX  
 XX 15-NOV-1994.  
 XX  
 XX 06-MAY-1993; 93JP-00105666.  
 XX  
 XX 06-MAY-1993; 93JP-00105666.  
 XX  
 XX 06-MAY-1993; 93JP-00105666.  
 XX  
 XX (KAENNO K.  
 XX (SUMO) SUMITOMO METAL IND LTD.  
 XX (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX  
 XX WPI, 1995-032330/05.  
 XX N-PSDB; AAQ80498.  
 XX  
 XX New HCV-originated proteinase active substance - used for site-specific  
 XX cleavage by an intermolecular reaction and the purification thereof.  
 XX  
 XX Disclosure; Page 10-19; 23pp; Japanese.  
 XX  
 XX This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is  
 XX cleaved between amino acids 2419 and 2420, by a new serine protease,  
 XX conng. the sequence of AAR86821. The proteinase is purified as a fused  
 XX product with the dihydrofolate reductase protein by using a methotrexate  
 XX column. It can be used for the development of an inhibitor for HCV  
 XX proteinase. (Updated on 16-OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 3010 AA;  
 XX  
 XX Query Match 95.6%; Score 1962; DB 2; Length 3010;  
 XX Best Local Similarity 93.6%; Pred. No. 9.4e-181;  
 XX Matches 368; Conservative 13; Mismatches 12; Indels 0; Gaps 0;  
 XX  
 XX 1 MAASCGGAVFTGLALTLSPYKYLRLIMWLQYLITRVEAHQVWVPLNVRGGRDAI 60  
 XX 814 MAASCGGAVFTGLALTLSPYKYLRLIMWLQYLITRVEAHQVWVPLNVRGGRDAI 873  
 XX  
 XX 61 ILTCAVPELIPITIKLLAIFGRLMVLQAGITKVFYFRAQGLIRACMLVRKAGHY 120  
 XX 874 ILTCAVPELIPITIKLLAIFGRLMVLQAGITKVFYFRAQGLIRACMLVRKAGHY 933  
 XX  
 XX 121 VQAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAAVEVPVPSDMEVKIITWADTAAC 180  
 XX 994 VQAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAAVEVPVPSDMEVKIITWADTAAC 993  
 XX  
 XX 181 GDIISGLPVSARRREILGPADNFEQGGWRLAPITAVSQOTRGLGCIITSITGRDN 240  
 XX 994 GDIISGLPVSARRREILGPADNFEQGGWRLAPITAVSQOTRGLGCIITSITGRDN 1053  
 XX  
 XX 241 QVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGKGPITQMTYTNVDQDLVQWA 300  
 XX 1054 QVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGKGPITQMTYTNVDQDLVQWA 1113  
 XX  
 XX 301 PPGARSMTPCTCGSSDLVYLRHADVPVRRRGSRSGLSPRVSYLKSSSGGFLCPS 360  
 XX 1114 PPGARSMTPCTCGSSDLVYLRHADVPVRRRGSRSGLSPRVSYLKSSSGGFLCPS 1173

QY	361	GHAVGIFPPAAVCTRGVAKAVDFIPVESHMETMR	393
Db	1174	GHVVGIFPPAAVCTRGVAKAVDFIPVESHMETMR	1206
RESULT 6			
AA68864			
ID	AA68864	standard; protein; 3010 AA.	
XX	AA68864;		
AC			
DT	06-DEC-1995	(first entry)	
XX			
DE	Hepatitis C virus RNA helicase.		
XX			
KW	Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;		
KW	hepatitis virus; recombinant production.		
XX			
OS	Hepatitis C virus.		
XX			
FH	Key	Location/Qualifiers	
FT	Region	196..198	
FT	Region	/label= N-linked glycosylation site	
FT	Region	209..211	
FT	Region	/label= N-linked glycosylation site	
FT	Region	234..236	
FT	Region	/label= N-linked glycosylation site	
FT	Region	250..252	
FT	Region	/label= N-linked glycosylation site	
FT	Region	305..307	
FT	Region	/label= N-linked glycosylation site	
FT	Region	325..327	
FT	Region	/label= N-linked glycosylation site	
FT	Region	417..419	
FT	Region	/label= N-linked glycosylation site	
FT	Region	423..425	
FT	Region	/label= N-linked glycosylation site	
FT	Region	430..432	
FT	Region	/label= N-linked glycosylation site	
FT	Region	448..450	
FT	Region	/label= N-linked glycosylation site	
FT	Region	532..534	
FT	Region	/label= N-linked glycosylation site	
FT	Region	556..558	
FT	Region	/label= N-linked glycosylation site	
FT	Region	576..578	
FT	Region	/label= N-linked glycosylation site	
FT	Region	623..625	
FT	Region	/label= N-linked glycosylation site	
FT	Region	645..647	
FT	Region	/label= N-linked glycosylation site	
FT	Region	1213..1215	
FT	Region	/label= N-linked glycosylation site	
FT	Region	1255..1257	
FT	Region	/label= N-linked glycosylation site	
FT	Region	2041..2043	
FT	Region	/label= N-linked glycosylation site	
FT	Region	2077..2079	
FT	Region	/label= N-linked glycosylation site	
FT	Region	2240..2242	
FT	Region	/label= N-linked glycosylation site	
FT	Region	2788..2790	
FT	Region	/label= N-linked glycosylation site	
XX			
PN	JP06319583-A.		
XX			
PD	22-NOV-1994.		
XX			
PF	18-SEP-1992;	92JP-00249241.	
XX			
PR	18-SEP-1992;	92JP-00249241.	

PA	(SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
XX	
DR	WPI; 1995-040330/06.
DR	N-PADB; AAQ81559.
XX	
PT	of hepatitis C virus helicase gene in baculovirus - useful for large
PT	scale prodn. of RNA helicase.
XX	
PB	Claim 1; Fig 1-4; 3pp; Japanese.
CC	
CC	AAQ81559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The DNA
CC	was used in the construction of an expression vector, which was used to
CC	transform a baculovirus host. The transformed baculovirus could then be
CC	used for the recombinant prodn. of HCV RNA helicase
XX	
SQ	Sequence 3010 AA;
Oy	Query Match            95.5%; Score 1961; DB 2; Length 3010; Best Local Similarity   93.6%; Pred. No. 1,2e+180; Matches 368; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
Dd	1 MAASCGAVFGLALLTLSPVKKLLRLIMWLQYLITRYEAHQVMIPPLNVRGGSDAI 60       814 MAASCGAVFGLALLTLSPVKKFLRLIMWLQYLITRYEAHLQVWPPLNVARGSDAI 873 
Oy	61 ILTCAVHELEPITKILLALFEGLPNVLQAGIRKVEFYFAQGILRACMLVRKAAGHY 120     874 ILLTCVAHELEPITKILLALIGLPLNLQAGITRVPEFYFAQGLIRACMLVRKAAGHY 933 
Oy	121 VQAAMFKLALTGYTYVDHTLPLODAHAAGRDLAVALVEPVIESDEVKIITGADTAAC 180     934 VQAAMFKLALTGYTYVDHTLPLEDMAHAGRDLAVALVEPVESDMEIKLIWGADTAAC 993 
Oy	181 GDITSGRPVSAARRGREILGPADFEGOGWRLLAPITAYVSQTGILGCIIITSLTGDKN 240     994 GDITSGRPVSARKREKELLGPAISFGSGWRLLAPIATYSQTGILGCIIITSLTGDKN 1053 
Dd	241 QVEGEVOVSTATOSFIATCNVGWCMTVFHGAGSKTLAPGXPTOMYTNYDDLVGNQA 300     1054 QVDEEVOLSTATOSFIATCNGVCWTIVYGAGSKTLAPGXAPTOMYTNYDDLVGMWA 1113 
Oy	301 PPGARSMPCTCCSSDIYLYTRHADVPYRRRGSRGSLSPPRSVTLKSSGGGPLICPS 360     1114 PPGARSMPCTCCSSDILYLYTRHADVPYRRRGDSRSSLSPRFTYTKSSGGGPLICPS 1173 
Oy	361 GHAVGIFFRAAVCTRGVAKAVDFIVESEMETMR 393     1174 GHVVGIFFRAAVCTRGVAKAVDFIVESEMETMR 1206 
Dd	
RESULT 7	
ID	ABG30601 standard; protein; 2201 AA.
XX	
AC	ABG30601;
DT	21-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX	
KM	Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX	cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.
OS	Hepatitis C virus.
OS	Synthetic.
FH	Key Location/Qualifiers
FT	Misc-difference 882 /label= Arg, Lys
FT	Misc-difference 2183 /note= "Wild type Met substituted by Thr"
XX	

PN W0200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 SQ Sequence 2201 AA;  
 XX  
 XX  
 Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7,1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MAASCGAVFTGLALTLSPYKVLARLWVLYITRVAHLQWIPPLNVRGGRDAI 60  
 DB 5 MAASCGAVFTGLALTLSPYKVLARLWVLYITRVAHLQWIPPLNVRGGRDAV 64  
 QY 61 ILTCAVHPELIPITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAGHY 120  
 DB 65 ILTCAHPELIPITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAGHY 124  
 QY 121 VQNAFKLALVTGYVDHLTPLODMAHAGLRDLAVALVEPVITSDNEVKITTWGADTAAC 180  
 DB 125 VQNALMMLALVTGYVDHLTPLODMAHAGLRDLAVALVEPVITSDNEVKITTWGADTAAC 184  
 QY 181 GDIISGLPVARGREITLGPADNFEQGRLLAPITAYVQQRGLGCTITSLTRGRDX 240  
 DB 185 GDITLGPVARGREITLGPADNFEQGRLLAPITAYVQQRGLGCTITSLTRGRDX 244  
 QY 241 QVEGEVQVSTATQSFATCVNGVQWTFVHAGSKTLAGEKGPITQWYTNVDDLVGWOA 300  
 DB 245 QVEGEVQVSTATQSFATCVNGVQWTFVHAGSKTLAGEKGPITQWYTNVDDLVGWOA 304  
 QY 301 PPGARSTPCTCGSSDLYVTRHADVI PVRRRDSGSLSPSPVYIKSGSGGLLCP 360  
 DB 305 PPGARSTPCTCGSSDLYVTRHADVI PVRRRDSGSLSPSPVYIKSGSGGLLCP 364  
 QY 361 GHAVGIFRAVCTRGVAKAVDFIPVBSMETTMR 393  
 DB 365 GHAVGIFRAVCTRGVAKAVDFIPVBSMETTMR 397

RESULT 8  
 ABG30591  
 ID ABG30591 standard; protein; 2201 AA.  
 XX  
 AC ABG30591;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 XX cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutetin.  
 XX  
 OS Hepatitis C virus.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 751 /note= "wild type Ser substituted by Gly"  
 FT FT Misc-difference 882 /label= Arg, Lys  
 FT FT  
 PN W0200252015-A2.  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 SQ Sequence 2201 AA;  
 XX  
 XX  
 Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7,1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MAASCGAVFTGLALTLSPYKVLARLWVLYITRVAHLQWIPPLNVRGGRDAI 60  
 DB 5 MAASCGAVFTGLALTLSPYKVLARLWVLYITRVAHLQWIPPLNVRGGRDAV 64  
 QY 61 ILTCAVHPELIPITKLLAIFGRLMVLQAGITKVPYFRAQGLIRACMLVRKAGHY 120







Db 245 QVEGEVQVSTATQSFATCVNGVCMVTHGAGSKTLGPKGITQMYTNVDQDLVGWQA 304  
QY 301 PPGARRMTPTCTGSSDLYLTRHADVIPIVRRRGRSGSLSPRPVSYLKSSGGPILCP8 360  
Db 305 PPGARRMTPTCTGSSDLYLTRHADVIPIVRRRGRSGSLSPRPVSYLKSSGGPILCP8 364  
QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPEMSETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDPFVPEMSETTMR 397

RESULT 12  
ABG30582  
ID ABG30582 standard; protein; 2201 AA.  
AC ABG30582;  
XX 21-OCT-2002 (first entry)  
XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.  
XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
XX cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
XX Hepatitis C virus.  
XX  
XX Key Location/Qualifiers  
XX Misc-difference 882 /note= "Wild type Lys substituted by Lys or Arg"  
XX Misc-difference 1233 /note= "Wild type Gly substituted by Cys"  
XX  
XX WO200252015-A2.  
XX 04-JUL-2002.  
XX 20-DEC-2001; 2001MO-CA001843.  
XX 22-DEC-2000; 2000US-0257857P.  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX Kukulj G, Pause A;  
XX WPI; 2002-575382/61.  
XX N-PSDB; ABK88574.  
XX  
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
XX possess enhanced transduction or replication efficiency, useful for  
XX evaluating potential inhibitors of HCV replication.  
XX  
XX Disclosure; Page 59-69; 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polypeptide  
XX region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence is encoded by the hepatitis C virus  
XX NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
XX sequence has been created from replicon APGK12 shown in ABG30581  
XX  
XX Sequence 2201 AA;  
XX  
XX Query Match 95.0%; Score 1951; DB 5; Length 2201;

Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;  
QY 1 MAASCGAVFICLALITTSPPYKVLARLIWLOVLTIRVRAH;QVWTPILNVGGPDAI 60  
Db 5 MAASCGAVFVGLITLTPSPHYKLPARLIWLOVLTIRVRAH;QVWTPILNVGGPDAV 64  
QY 61 ILITCAVHELIFFDITKLLAIFGPLYMLOAGITVVPFVBAQGLIRACMLVRKAAGHY 120  
Db 65 ILITCAIHELIFFITKLLAIFGPLYMLOAGITVVPFVBAQGLIRACMLVRKAAGHY 124  
QY 121 VQMAFMKLAALTGTIVYDHLTPLODMANAGLDLAVANPEVTFESMEVKIITWGDITAC 180  
Db 125 VQMALMKLAALTGTIVYDHLTPLODMANAGLDLAVANPEVTFESMEVKIITWGDITAC 184  
QY 181 GDIIISGLPVASARRGEIILGPADNFEQGWRLIAPITVYASQOTRGLGCIITSLTGRDN 240  
Db 185 GDIIISGLPVASARRGEIILGPADNFEQGWRLIAPITVYASQOTRGLGCIITSLTGRDN 244  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTHGAGSKTLGPKGITQMYTNVDQDLVGWQA 300  
Db 245 QVEGEVQVSTATQSFATCVNGVCMVTHGAGSKTLGPKGITQMYTNVDQDLVGWQA 304  
QY 301 PPGARRMTPTCTGSSDLYLTRHADVIPIVRRRGRSGSLSPRPVSYLKSSGGPILCP8 360  
Db 305 PPGARRMTPTCTGSSDLYLTRHADVIPIVRRRGRSGSLSPRPVSYLKSSGGPILCP8 364  
QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPEMSETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDPFVPEMSETTMR 397

RESULT 13  
ABG30580  
ID ABG30580 standard; protein; 2201 AA.  
AC ABG30580;  
XX 21-OCT-2002 (first entry)  
XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.  
XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
XX Hepatitis C virus.  
XX  
XX Key Location/Qualifiers  
XX Misc-difference 882 /note= "Encoded by ARG"  
XX  
XX WO200252015-A2.  
XX 04-JUL-2002.  
XX 20-DEC-2001; 2001MO-CA001843.  
XX 22-DEC-2000; 2000US-0257857P.  
XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX Kukulj G, Pause A;  
XX WPI; 2002-575382/61.  
XX  
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
XX possess enhanced transduction or replication efficiency, useful for  
XX evaluating potential inhibitors of HCV replication.  
XX  
XX Disclosure; Page 69-74; 140pp; English.  
XX  
XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX

CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;

Best Local Similarity 93.1%; Pred. No. 7.1e-180; Mismatches 13; Indels 0; Gaps 0;

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFVIGLALTLSPYKVLARLIMLQYLITREVAHLQVWIPPLNVRGSDAI 60  
 DB 5 MAASCGAVFVIGLALTLSPYKVLARLIMLQYLITREVAHLQVWIPPLNVRGSDAV 64  
 QY 61 ILITCAVHPELIFDITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
 DB 65 ILITCAHPELIFITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 124  
 QY 121 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPVFSDMEVKIITWGAADTAAC 180  
 DB 125 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPVFSDMEVKIITWGAADTAAC 184  
 QY 181 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQQTRGLGCIITSLTGRDN 240  
 DB 185 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQQTRGLGCIITSLTGRDN 244  
 QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAQRKGPITQMTYTNVDDLVGMOA 300  
 DB 245 QVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAQRKGPITQMTYTNVDDLVGMOA 304  
 QY 301 PPGARSMTPCTCGSSDLYLVTTRADVI PVRRRGRDSSGSLSPRVSYLKSSGGPILCP 360  
 DB 305 PPGARSLTPCTCGSSDLYLVTTRADVI PVRRRGRDSSGSLSPRVSYLKSSGGPILCP 364  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
 DB 365 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 397

RESULT 14

ABG30587  
 ID ABG30587 standard; protein; 2201 AA.

XX AC ABG30587;  
 XX DT 21-OCT-2002 (first entry)  
 XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #7.  
 XX KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 XX KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 XX OS Hepatitis C virus.  
 XX PN WO200252015-A2.  
 XX PD 04-JUL-2002.  
 XX PF 20-DEC-2001; 2001MO-CA001843.  
 XX PR 22-DEC-2000; 2000US-0257857P.  
 XX PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;  
 PI WPI, 2002-575382/61.  
 DR N-PsDr; ABK88587.  
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 PS Disclosure; Page 120-129; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA;

Query Match 95.0%; Score 1951; DB 5; Length 2201;

Best Local Similarity 93.1%; Pred. No. 7.1e-180; Mismatches 13; Indels 0; Gaps 0;

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFVIGLALTLSPYKVLARLIMLQYLITREVAHLQVWIPPLNVRGSDAI 60  
 DB 5 MAASCGAVFVIGLALTLSPYKVLARLIMLQYLITREVAHLQVWIPPLNVRGSDAV 64  
 QY 61 ILITCAVHPELIFDITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 120  
 DB 65 ILITCAHPELIFITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAAGHY 124  
 QY 121 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPVFSDMEVKIITWGAADTAAC 180  
 DB 125 VQMAFMKLAALTGYYVDHITPLQDMAHAGRLDAVAVEPVFSDMEVKIITWGAADTAAC 184  
 QY 181 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQQTRGLGCIITSLTGRDN 240  
 DB 185 GDIIISGLPVSARRGREIILGPADNFEQGWRLAPITAYSQQTRGLGCIITSLTGRDN 244  
 QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAQRKGPITQMTYTNVDDLVGMOA 300  
 DB 245 QVEGEVQVSTATQSFATCVNGVCMVTFHGAGSKTLAQRKGPITQMTYTNVDDLVGMOA 304  
 QY 301 PPGARSMTPCTCGSSDLYLVTTRADVI PVRRRGRDSSGSLSPRVSYLKSSGGPILCP 360  
 DB 305 PPGARSLTPCTCGSSDLYLVTTRADVI PVRRRGRDSSGSLSPRVSYLKSSGGPILCP 364  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
 DB 365 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 397

RESULT 15

ABG30599  
 ID ABG30599 standard; protein; 2201 AA.

XX AC ABG30599;  
 XX DT 21-OCT-2002 (first entry)  
 XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #8.

KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 892 /label= Arg, Lys  
 FT Misc-difference 1346 /note= "wild type leu substituted by pro"  
 FT  
 XX WO200252015-A2.  
 PN 04-JUL-2002.  
 PD  
 XX 20-DEC-2001; 2001WO-CA001843.  
 PF  
 XX 22-DEC-2000; 2000US-0257857P.  
 PR  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA  
 XX Kukulj G, Pause A;  
 PI  
 XX WPI; 2002-575382/61.  
 DR  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 PS  
 XX Claim 3; Page; 140pp; English.  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APGX12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX  
 XX Sequence 2201 AA;  
 SQ

Query Match 95.0%; Score 1951; DB 5; Length 2201;  
 Best Local Similarity 93.1%; Pred. No. 7.1e-180;  
 Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGGAVFGLALLTSPYKVLARLIMWLQYLITRVAHLQVWIPPLNVRGGDAI 60  
 DB 5 MAASCGGAVFGLALLTSPYKVLARLIMWLQYLITRVAHLQVWIPPLNVRGGDAV 64  
 QY 61 ILTCAHPELIPITITLLAIFGPMVQAGITKVPYFRAOGLIRACMLVRKAGGHY 120  
 DB 65 ILTCAHPELIPITITLLAIFGPMVQAGITKVPYFRAOGLIRACMLVRKAGGHY 124  
 QY 121 VQNAFMLAALTGTYYVDHTPLQDMWAGLRDLAAVEVIFSDMEVKIITWGAADTAAC 180  
 DB 125 VQNAFMLAALTGTYYVDHTPLQDMWAGLRDLAAVEVIFSDMEVKIITWGAADTAAC 184  
 QY 181 GDIISGLPVGARRREILLPADNFEQGWRLAPITAYSQTRGLIGCITISLTGRDKY 240  
 DB 185 GDIISGLPVGARRREILLPADNFEQGWRLAPITAYSQTRGLIGCITISLTGRDRY 244

QY 241 QVEGEVQVNSTATOSFLATCVNGVCMVTFHAGSKTLTAGPKPITOMYTNVDQDLVGMOA 300  
 DB 245 QVEGEVQVNSTATOSFLATCVNGVCMVTFHAGSKTLTAGPKPITOMYTNVDQDLVGMOA 304  
 QY 301 PEGARSMTPTCTGSSDLYLVRHADVIPRRRGRSGSLSPRVSYLKSSGGPPLCP 360  
 DB 305 PEGARSMTPTCTGSSDLYLVRHADVIPRRRGRSGSLSPRVSYLKSSGGPPLCP 364  
 QY 361 GHAVGIFRAAVCTRGVAKAVDPFIPVSMETMR 393  
 DB 365 GHAVGIFRAAVCTRGVAKAVDPFIPVSMETMR 397

Search completed: May 6, 2004, 09:30:45  
 Job time : 54.493 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 12.7992 Seconds  
(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-11

Perfect score: 2053  
Sequence: 1 MAASCGAVFIGLALITLSP.....RGVAKAVDPFVPSMETMTR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: PIR\_78:\*  
2: pir1:\*  
3: pir2:\*  
4: pir3:\*  
5: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1967	95.8	3010	1 GNMVCU	genome polypeptide
2	1950	95.0	3010	1 A45573	genome polypeptide
3	1935	94.3	3010	1 GNMVTM	genome polypeptide
4	1910	93.0	3010	1 S18030	genome polypeptide
5	1888	92.0	3010	1 GNMVTC	genome polypeptide
6	1764	85.9	3011	1 GNMWC3	genome polypeptide
7	1758	85.6	3011	1 S40770	genome polypeptide
8	1752	85.3	3011	1 GNMVCH	genome polypeptide
9	1493	72.7	3014	1 JCS620	genome polypeptide
10	1403	68.3	3033	1 J01303	genome polypeptide
11	1401	68.2	3033	1 GNMVU8	genome polypeptide
12	408.5	19.9	3005	2 T08841	polypeptide - dour
13	342.5	16.7	2970	2 T08839	polypeptide - marm
14	112	5.5	692	2 H71426	hypothetical prote
15	102.5	5.0	660	2 VHMWH2	structural protein
16	101.5	4.9	3434	1 GNMVWV	genome polypeptide
17	101	4.9	564	2 S36637	signal recognition
18	101	4.9	600	2 S46642	DNA-directed DNA p
19	100.5	4.9	353	2 G87352	conserved hypotet
20	99	4.8	399	2 AH3038	conserved hypotet
21	99	4.8	399	2 C98247	hypothetical 50.8k
22	99	4.8	446	2 AF1509	conserved hypotet
23	99	4.8	451	2 H82044	C4-dicarboxylate t
24	97.5	4.7	1085	2 T03531	cohn protein homol
25	95.5	4.7	470	2 JC4098	tetracycline 6-hyd
26	94.5	4.6	904	2 A84212	hypothetical prote
27	94.5	4.6	2796	2 JC4743	fatty-acid synthas
28	94	4.6	434	2 G82728	conserved hypotet
29	93.5	4.6	418	2 H90679	probable transport

30	93.5	4.6	418	2 D85530	probable transport
31	93.5	4.6	477	2 H75026	oligopeptide abc t
32	93.5	4.6	1380	2 T18309	receptor-adenylate
33	93.5	4.6	3069	2 H70656	fatty-acid synthas
34	93	4.5	3414	1 GNMVNE	genome polypeptide
35	93	4.5	7463	2 T36248	CDA peptide synthe
36	92.5	4.5	665	2 D83252	nucleotide sugar e
37	92.5	4.5	706	2 S33761	transferrin precu
38	92.5	4.5	716	2 G83612	hypothetical prote
39	92	4.5	659	1 B44212	structural protein
40	91.5	4.5	401	1 A36961	pillin biogenesis p
41	91.5	4.5	428	2 AF0241	probable coenzyme
42	91.5	4.5	446	2 A31150	conserved hypotet
43	91.5	4.5	3412	1 GNMVTE	genome polypeptide
44	90.5	4.4	418	2 A64763	probable transport
45	90.5	4.4	868	2 H81775	aconitate hydratase

## ALIGNMENTS

### RESULT 1

GNMVCU genome polypeptide - hepatitis C virus (strain J)

N:contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Jun-1992 #sequence\_revision 30--Jun-1992 #text\_change 19-Jun-2001

C/Accession: A39253; PS0086 R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients v

A:Reference number: A39253; PMID:91088550; PMID:2175903

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D90208; NID:G221610; PID:BA1423.1; PID:G221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KA2>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polypeptide have not been determined.

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine

F:2-115/Product: capsid protein C #status predicted <CPC>

F:16-191/Product: envelope protein M #status predicted <EM>

F:192-389/Product: major envelope protein E #status predicted <MEB>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitisin #status predicted <NS3>

F:2230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NA>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NB>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,233,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2;

Query Match 95.8%; Score 1967; DB 1; Length 3010;

Best Local Similarity 93.9%; Pred. No. 1.8e-153;

Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

QY	1 MAASCGAVFIGLALITLSPYKVLARLIMWLQYLIRVENHLOWVPIPLNVRGGRDAI 60
DB	814 MAASCGAVFVGLVLTISPYKVFILALIMWLQYFIRAEHLOWVPIPLNVRGGRDAI 873
QY	61 ILLTCAVPELIFDTIKLLAIFGPLYVLQAGITKVPYFVRAQGLIRACMLVRKAGGHY 120
DB	874 ILLTCAVPELIFDTIKLLAIFGPLYVLQAGITKVPYFVRAQGLIRACMLVRKAGGHY 933

QY 121 VQMAFKMLAALTGTYVDHLTPLODMAHAGRLDAVAEVPFISDMEVKITITMGADTAAC 180  
 DB 934 VQMAFKMLAALTGTYVDHLTPLODMAHAGRLDAVAEVPFISDMEVKITITMGADTAAC 993  
 QY 181 GDIISGLPVSARRRREILGPDADNFEQGWRLAPITAVSQQRGLLGLITSLTGRDKN 240  
 DB 994 GDIISGLPVSARRRREILGPDADNFEQGWRLAPITAVSQQRGLLGLITSLTGRDKN 1053  
 QY 241 QVEGEVQVVSATQSPFLATCNGVCMWTFPHAGSKTLAGEKGPITQMTYNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVVSATQSPFLATCNGVCMWTFPHAGSKTLAGEKGPITQMTYNVDQDLVGMQA 1113  
 QY 301 PPGARSTPCTCGSSDLYLTRHADVPVRRRGDSRLSPRVSTLTKSSGGPILCP 360  
 DB 1114 PPGARSTPCTCGSSDLYLTRHADVPVRRRGDSRLSPRVSTLTKSSGGPILCP 1173  
 QY 361 GHAVGIFRAVACTRGVAKAVDFIPVESMETTMR 393  
 DB 1174 GHAVGIFRAVACTRGVAKAVDFIPVESMETTMR 1206

## RESULT 2

A45573  
 genome polyprotein - hepatitis C virus (strain J7)  
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: A45573  
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,  
 Virus Res. 23, 39-53, 1992  
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  
 A:Reference number: A45573; MUID:9225714; PMID:1318627  
 A:Accession: A45573  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3010 <TAN>  
 A:Cross-references: GB:D11168; GB:D01171; NID:9221612; PID:BA01943.1; PID:9221613  
 A:Experimental source: HCV-JT  
 A:Note: Sequence extracted from NCBI backbone (NCIN:106206, NCBI:106207)  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Product: hepatitis virus #status predicted <NS3>  
 F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F:1316-1319/Region: nucleotide-binding motif B  
 F:1316-1862/Product: nonstructural protein NS4 #status predicted <N4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 95.0%; Score 1950; DB 1; Length 3010;

Best Local Similarity 93.6%; Pred. No. 4.7e-152;

Matches 369; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFVGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGHY 120  
 DB 874 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGHY 933  
 QY 121 VQMAFKMLAALTGTYVDHLTPLODMAHAGRLDAVAEVPFISDMEVKITITMGADTAAC 180  
 DB 934 VQMAFKMLAALTGTYVDHLTPLODMAHAGRLDAVAEVPFISDMEVKITITMGADTAAC 993  
 QY 181 GDIISGLPVSARRRREILGPDADNFEQGWRLAPITAVSQQRGLLGLITSLTGRDKN 240

DB 994 GDIISGLPVSARRRREILGPDADNFEQGWRLAPITAVSQQRGLLGLITSLTGRDKN 1053  
 QY 241 QVEGEVQVVSATQSPFLATCNGVCMWTFPHAGSKTLAGEKGPITQMTYNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVVSATQSPFLATCNGVCMWTFPHAGSKTLAGEKGPITQMTYNVDQDLVGMQA 1113  
 QY 301 PPGARSTPCTCGSSDLYLTRHADVPVRRRGDSRLSPRVSTLTKSSGGPILCP 360  
 DB 1114 PPGARSTPCTCGSSDLYLTRHADVPVRRRGDSRLSPRVSTLTKSSGGPILCP 1173  
 QY 361 GHAVGIFRAVACTRGVAKAVDFIPVESMETTMR 393  
 DB 1174 GHAVGIFRAVACTRGVAKAVDFIPVESMETTMR 1206

## RESULT 3

GNMVTM  
 genome polyprotein - hepatitis C virus (strain Taiwan)  
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A40244  
 R:Chen, P.T.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
 Virology 188, 102-113, 1992  
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A:Reference number: A40244; MUID:92230206; PMID:1314449  
 A:Accession: A40244  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <CHE>  
 A:Cross-references: GB:M64754  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <N4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,223,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 94.3%; Score 1935; DB 1; Length 3010;

Best Local Similarity 91.9%; Pred. No. 8.1e-151;

Matches 361; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFVGLALTLSPYKVLARLIMLOYLITRVEAHQWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGHY 120  
 DB 874 ILITCAVPELIPDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGHY 933  
 QY 121 VQMAFKMLAALTGTYVDHLTPLODMAHAGRLDAVAEVPFISDMEVKITITMGADTAAC 180  
 DB 934 VQMAFKMLAALTGTYVDHLTPLODMAHAGRLDAVAEVPFISDMEVKITITMGADTAAC 993  
 QY 181 GDIISGLPVSARRRREILGPDADNFEQGWRLAPITAVSQQRGLLGLITSLTGRDKN 240  
 DB 994 GDIISGLPVSARRRREILGPDADNFEQGWRLAPITAVSQQRGLLGLITSLTGRDKN 1053  
 QY 241 QVEGEVQVVSATQSPFLATCNGVCMWTFPHAGSKTLAGEKGPITQMTYNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVVSATQSPFLATCNGVCMWTFPHAGSKTLAGEKGPITQMTYNVDQDLVGMQA 1113

QY 301 PGARSMPTCTCGSSDLYLVTRHADVT PVRRRGDSRGSLSRPVSYLKSSGGPILCPs 360  
 Db 1114 PGARSMPTCTCGSSDLYLVTRHADVT PVRRRGDSRGSLSRPVSYLKSSGGPILCPs 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393  
 Db 1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 1206

## RESULT 4

S18030  
 genome polypeptide - hepatitis C virus (isolate JX1)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Variety: isolate JX1  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
 C:Accession: S18030; S35570; A48332; S18029  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 Submitted to the EMBL Data Library, September 1991  
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
 A:Reference number: S18028  
 A:Accession: S18030  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <HON>  
 A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479  
 A:Experimental source: isolate JX1 from an individual  
 R:Honda, M.; Kaneko, S.; Ueno, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 128, 163-169, 1993  
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A:Reference number: A48332; MUID:93119270; PMID:8380322  
 A:Accession: S33570  
 A:Molecule type: genomic RNA  
 A:Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'D', 655-761, 'T', 763-782 <HON>  
 A:Cross-references: EMBL:X61591  
 A:Note: this sequence is inconsistent with the nucleotide translation  
 A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
 A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 93.0%; Score 1910; DB 1; Length 3010;  
 Best Local Similarity 91.8%; Pred. No. 9.3e-149;  
 Matches 361; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 1 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHQVWIPPLNVRGRDAI 60  
 Db 814 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHQVWIPPLNVRGRDAI 873  
 QY 61 ILTCAVHPELFDITKLLAIFGLMVLQAGITKVPYFAAGLIRACMLVRKAAGHY 120  
 Db 874 ILTCAVHPELFDITKLLAIFGLMVLQAGITKVPYFAAGLIRACMLVRKAAGHY 933  
 QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITWGADTAAC 180  
 Db 934 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITWGADTAAC 993  
 QY 181 GDIIIGLPEVSARGRBELLGPADNFBEGQWLLAPITVYSCQTRGLGCIITISLGRDXN 240

Db 994 GDIIIGLPEVSARGRBELLGPADNFBEGQWLLAPITVYSCQTRGLGCIITISLGRDXN 1053  
 QY 241 QVEGEVQVSTATQGFPLATCVNGVCWTFVHGAGSKTLAGPFGPTOMTYNDQDILVQQA 300  
 Db 1054 QVEGEVQVSTATQGFPLATCVNGVCWTFVHGAGSKTLAGPFGPTOMTYNDQDILVQQA 1113  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVT PVRRRGDSRGSLSRPVSYLKSSGGPILCPs 360  
 Db 1114 PGARSMPTCTCGSSDLYLVTRHADVT PVRRRGDSRGSLSRPVSYLKSSGGPILCPs 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393  
 Db 1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 1206

## RESULT 5

GNVATC  
 genome polypeptide - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
 C:Accession: A38465  
 R:Takamizawa, A.; Mori, C.; Fukey, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
 A:Title: Structure and organization of the hepatitis C virus genome isolated from human  
 A:Reference number: A38465; MUID:91140698; PMID:1847440  
 A:Accession: A38465  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <TKX>  
 A:Cross-references: EMBL:M58335; NID:G329770; PIDN:AA72945.1; PID:G329771  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 92.0%; Score 1888; DB 1; Length 3010;  
 Best Local Similarity 90.8%; Pred. No. 6.1e-147;  
 Matches 357; Conservative 13; Mismatches 23; Indels 0; Gaps 0;

QY 1 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHQVWIPPLNVRGRDAI 60  
 Db 814 MAASCGAVFICGLALTLSPYKVLARLIWVLOYLIRVEAHQVWIPPLNVRGRDAI 873  
 QY 61 ILTCAVHPELFDITKLLAIFGLMVLQAGITKVPYFAAGLIRACMLVRKAAGHY 120  
 Db 874 ILTCAVHPELFDITKLLAIFGLMVLQAGITKVPYFAAGLIRACMLVRKAAGHY 933  
 QY 121 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITWGADTAAC 180  
 Db 934 VQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITWGADTAAC 993  
 QY 181 GDIIIGLPEVSARGRBELLGPADNFBEGQWLLAPITVYSCQTRGLGCIITISLGRDXN 240  
 Db 994 GDIIIGLPEVSARGRBELLGPADNFBEGQWLLAPITVYSCQTRGLGCIITISLGRDXN 1053  
 QY 241 QVEGEVQVSTATQGFPLATCVNGVCWTFVHGAGSKTLAGPFGPTOMTYNDQDILVQQA 300  
 Db 1054 QVEGEVQVSTATQGFPLATCVNGVCWTFVHGAGSKTLAGPFGPTOMTYNDQDILVQQA 1113  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVT PVRRRGDSRGSLSRPVSYLKSSGGPILCPs 360

Dd	1114	PFGARSLTPTCCGSSDLYLTRADVPVRRGRDSRSLSPRVSYLKSGSGPILCPF	1173
Qy	361	GHAAGIFRAAYCTRGVAKAVDFIPVESMETMR	393
Dd	1174	GHAAGIFRAAYCTRGVAKAVDFIPVESMETMR	1206

## RESULT

genome polyprotein - hepatitis C virus (strain HCV-1)  
N, contains capsid protein C, envelope protein M, hepatitis A (EC 3.4.21.98) (nonstructural protein NS4a, nonstructural protein NS4b, nonstructural protein NS5  
C, species: hepatitis C virus  
C, Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text change 19-Jan-2001  
C, Accession: A39166; PQ0403; PQ0404  
R, Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coatsworth, J.E., Rojkind, M., Krawczynski, K., Holmes, E.C., Dow, B., Feunteun, J.F., Follett, E., Yap, P.L., Gen. Virol. 73, 1131-1144, 1992  
A, Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A, Reference number: PQ0393; PMID:92268871; PMID:1316939  
A, Accession: PQ0403  
A, Molecule type: genomic RNA  
A, Residues: 1577-1633 <CHA>  
A, Cross-references: DDBJ:D10128  
A, Experimental source: Isolates E-b16  
A, Accession: PQ0404  
A, Status: preliminary  
A, Molecule type: genomic RNA  
A, Residues: 1577-1633 <CH>  
A, Experimental source: Isolates E-b16  
A, Status: preliminary  
A, Molecule type: genomic RNA  
A, Residues: 1577-1633 <CH>  
A, Experimental source: Isolates E-b17  
C, Superfamily: hepatitis C virus genome polyprotein  
C, Keywords: ATP; capsid protein C, envelope protein; glycoprotein; hydrolase; nonstructural  
R, 1115/Product: capsid protein C #status predicted <GPC>  
F, 1115-151/Product: envelope protein M #status predicted <BPM>  
F, 1195-389/Product: major envelope protein E #status predicted <MPE>  
F, 1390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F, 1730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F, 1007-1615/Product: hepatitis A #status predicted <NS3>  
F, 1230-1237/Region: nucleotide-binding motif A (P-loop)  
F, 1312-1317/Region: nucleotide-binding motif B  
F, 1316-1319/Region: DEXH motif  
F, 1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F, 1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F, 2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F, 1196, 209, 224, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 22

Query Match	85.9%	Score 1764	DB 1	Length 3011
Best Local Similarity	81.9%	Pred. No. 1e-136		
Matches	322	Conservative	34	Mismatches 37, Indels 0, Gaps 0

  

QY	1	MAASCGAVFTGLALITLSPYKVLARLIMWLQYLITRVEAHLQWVPEPLNVEGGDPAI	60
		.....:.....	
DB	814	VAASGCGVVLVGLMALITLSPYKRYISWCMWLQYFLTRVEAQLHWVPEPLNVEGGRAV	873
QY	61	ILITCAVPEPLFDITKLLIAIFGFLMWLOAGITVPEPYVAOGLIRACMLVRAAGGHY	120
		.....:.....	
DB	874	ILIMCAVPEPLFDITKLLIAVFGFLMWLOASLKVPEPYVQGLIRCALARKMIGHY	933
QY	121	VQMAFMKLAALTGIVYVDHLTPLODMAHAGLRDLAAVEPVIIFSDMEVXKITMGADTAAC	180
		.....:.....	
DB	934	VQWVIRIKGALTGTIVYVNHLPRLRDWANGRLDLAAVEAPVFSMEKLTITMGADTAAC	993
QY	181	GDIIISGFPVSRARGEIILGPADNVEGGWMLLIATVYSOOTRSLGCIITSLTGRPXN	240
		.....:.....	
DB	994	GDIIINGLVSARGEIILGPADGVSQMKMLLIATVYAQOTRGLGCIITSLTGRPXN	1053

[illegible]

## RESULT 7

genome polyprotein - hepatitis C virus  
 NCContains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
 CSpecies: hepatitis C virus  
 CDate: 19-May-2000 #sequence\_revision 19-May-2000 #ext\_change 19-Jan-2001  
 CAccession: S40770; PCl285  
 ROkamoto, H.  
 submitted to the EMBL Data Library, March 1992  
 AReference number: S40770  
 AAccession: S40770  
 AMolecule type: genomic RNA  
 AResidues: 1-3011 <OK>  
 ACross-references: EMBL:DI0749; NID:G221586; PIDN:BA001582.1; PID:G221587  
 ROkamoto, H.; Okeda, S.; Sugiyama, Y.; Yotsunoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
 ATitle: The 5'-terminal sequence of the hepatitis C virus genome.  
 AReference number: PCl284; MUID:9103116; PMID:2170712  
 AAccession: PCl285  
 AMolecule type: genomic RNA  
 AResidues: 1-513 <OK>  
 ACross-references: GB:000831; NID:G221511; PIDN:BA00705.1; PID:G221512  
 AExperimental source: isolate HC-J1  
 CSuperfamily: hepatitis C virus genome polyprotein  
 CKeywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
 F12-115/Product: capsid protein C #status predicted <CPC>  
 F116-191/Product: envelope protein M #status predicted <EPM>  
 F132-389/Product: major envelope protein E #status predicted <NEE>  
 F132-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F1007-1615/Product: hepacivirin #status predicted <NS3>  
 F1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F1312-1317/Region: nucleotide-binding motif B  
 F1316-1319/Region: DEXH motif  
 F1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>  
 F1863-2013/Product: nonstructural protein NS4b #status predicted <NA5>  
 F2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

[illegible]



```

Db      1054 QVEGEVQIVSTATQTFELATCINGVCMTVYHGAGRTTASPKGPIQITNTVDOLVGNPA 1113
QY      301 PGARSMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPVSYLKSSGGPILCPG 360
Db      1114 PGARSMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPVSYLKSSGGPILCPG 1173
QY      361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 393
Db      1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

```

## RESULT 8

```

GENWYCH
genome polypeptide - hepatitis C virus (strain H)
N:contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
protein NS4b; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Accession: A36814; A41546
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
R:Inchuspe, G.; Zebdeed, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Submitted to Genbank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <N>
A:Cross-references: GB:M67463; NID:G329737; PIND:AAA45534.1; PID:G329738
R:Inchuspe, G.; Zebdeed, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructu
F:1.115/Product: capsid protein C #status predicted <CPC>
F:1.115/Product: envelope protein M #status predicted <EMP>
F:1.191/Product: major envelope protein E #status predicted <MEB>
F:1.389/Product: nonstructural protein NS1 #status predicted <NS1>
F:1.390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F:1.707-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F:1.1007-1616/Product: hepatitis C virus predicted <NS4>
F:1.130-1237/Product: nucleotide-binding motif A (P-loop)
F:1.131-1317/Product: nucleotide-binding motif B
F:1.131-1317/Product: DEHX motif
F:1.131-1317/Product: nonstructural protein NS4 #status predicted <NS4>
F:1.161-1863/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1.163-2011/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,334,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23
F:196,209,334,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

```

```

Query Match      85.3%; Score 1752; DB 1; Length 3011;
Best Local Similarity 81.4%; Pred. No. 1e-135;
Matches 320; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

```

```

QY      1 MAASCGAVFTGLTLTSPYKYLARLIMWLOYLTVREAHQWIPPLNVRGSDAI 60
Db      814 VAAACGGVTLVGLTALTLSPYKYLARLIMWLOYLTVREAHQWIPPLNVRGSDAI 873
QY      61 ILTCAVHPELIPITIKLLAIFGLPWLVAQITKVEFYVAQGLIRACMLVRKAAGHY 120
Db      874 ILTCAVHPELIPITIKLLAIFGLPWLVAQITKVEFYVAQGLIRACMLVRKAAGHY 933
QY      121 VQAPFMTALATGYVVDHLTPLODMNAGRLDVAVEPIFEDMEVKIITWADPTAAC 180
Db      934 VQAPFMTALATGYVVDHLTPLODMNAGRLDVAVEPIFEDMEVKIITWADPTAAC 993
QY      181 GDIISGLPVSARREILIGPADNFEQGGWLLAPITAYSOQTGLGCIITSLTGKDN 240
Db      994 GDIISGLPVSARREILIGPADNFEQGGWLLAPITAYSOQTGLGCIITSLTGKDN 1053
QY      241 QVEGEVQIVSTATQTFELATCINGVCMTVYHGAGRTTASPKGPIQITNTVDOLVGNPA 300

```

```

Db      1054 QVEGEVQIVSTATQTFELATCINGVCMTVYHGAGRTTASPKGPIQITNTVDOLVGNPA 1113
QY      301 PGARSMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPVSYLKSSGGPILCPG 360
Db      1114 PGARSMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPVSYLKSSGGPILCPG 1173
QY      361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 393
Db      1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

```

## RESULT 9

```

JCS620
genome polypeptide - hepatitis C virus (isolate EUH1480)
N:contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
protein NS4b; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Accession: JCS620
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A:Reference number: JCS620; MUID:9736593; PMID:9223423
A:Accession: JCS620
A:Molecule type: mRNA
A:Residues: 1-3014 <CNA>
A:Cross-references: GB:Y13184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
F:1.115/Product: capsid protein C #status predicted <CPC>
F:1.191/Product: envelope protein M #status predicted <EMP>
F:1.389/Product: major envelope protein E #status predicted <MEB>
F:1.390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:1.707-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1.1007-1616/Product: hepatitis C virus predicted <NS4>
F:1.130-1237/Product: nucleotide-binding motif A (P-loop)
F:1.131-1317/Product: nucleotide-binding motif B
F:1.131-1317/Product: DEHX motif
F:1.161-1863/Product: nonstructural protein NS4 #status predicted <NS4>
F:1.163-2011/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Product: interferon sensitivity determining #status predicted

```

```

Query Match      72.7%; Score 1493; DB 1; Length 3014;
Best Local Similarity 68.8%; Pred. No. 2.3e-114;
Matches 267; Conservative 56; Mismatches 65; Indels 0; Gaps 0;

```

```

QY      6 GCAVFTGLTLTSPYKYLARLIMWLOYLTVREAHQWIPPLNVRGSDAI 65
Db      820 GCAVFTGLTLTSPYKYLARLIMWLOYLTVREAHQWIPPLNVRGSDAI 879
QY      66 AVHPELIPITIKLLAIFGLPWLVAQITKVEFYVAQGLIRACMLVRKAAGHYQNAF 125
Db      880 AVHPELIPITIKLLAIFGLPWLVAQITKVEFYVAQGLIRACMLVRKAAGHYQNAF 939
QY      126 MKLAALATGYVVDHLTPLODMNAGRLDVAVEPIFEDMEVKIITWADPTAAC 185
Db      940 MKLAALATGYVVDHLTPLODMNAGRLDVAVEPIFEDMEVKIITWADPTAAC 999
QY      186 GDIISGLPVSARREILIGPADNFEQGGWLLAPITAYSOQTGLGCIITSLTGKDN 245
Db      1000 GDIISGLPVSARREILIGPADNFEQGGWLLAPITAYSOQTGLGCIITSLTGKDN 1059
QY      246 VQVSTATQTFELATCINGVCMTVYHGAGRTTASPKGPIQITNTVDOLVGNPA 305
Db      1060 VQVSTATQTFELATCINGVCMTVYHGAGRTTASPKGPIQITNTVDOLVGNPA 1119
QY      306 SMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPVSYLKSSGGPILCPGSHAG 365
Db      1120 SMTPTCTGSSDLYLVTRHADVPVRRGDSRGSLSPPVSYLKSSGGPILCPGSHAG 1179

```

366 IFRAAVCTRGVAKAVDFIPVSMETMR 393  
1180 VFRAAVCTRGVAKLREVPVENLETMR 1207

RESULT 10

genome polyprotein - hepatitis C virus (isolate HC-J6)  
N:contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C:Species: hepatitis C virus  
C:Accession: J01303  
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 17-Nov-2000  
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
A:Reference number: J01303; MUID:9204440; PMID:1658136  
A:Accession: J01303  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651  
A:Experimental source: isolate HC-J6 from a Japanese individual  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polyprotein; serine proteinase; transmembrane; envelope protein M #status predicted <CPC>  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <CPC>  
F:116-191/Product: major envelope protein E #status predicted <MEE>  
F:330-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1667-2017/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 68.3%; Score 1403; DB 1; Length 3033;  
Best Local Similarity 64.9%; Pred. No. 6,2e-107;  
Matches 252; Conservative 58; Mismatches 78; Indels 0; Gaps 0;

6 GGAVFGLALLTSPYKVLARLIMLQYLITRVEAHLOVWIPNLVNRGRDAIILITLC 65  
823 GAAVLVITLFTLPGYKTLISRFMLCYLLLAEMVQEMAPMVQVGRGCIIMAV 882  
66 AVHDELIFDITKLLAIFGLMLVQAGITVYFVRAQGLIRACMLVRKAAGHYVMAR 125  
883 IFGGVVFEDITKMLLAIVLGPAYLLKGLITREVPFVRAHALRMTVVRHLAGRYYQVNL 942  
126 MKLAALGTIVYDHLTPLODMAHAGLRDLAVAVEPVFSDMEVKIITWGDPTAACGDIIS 185  
943 IALGRMTGTYIDHLTPMSMANGLRDLAVAVEPVFSDMEVKIITWGEETIACGDIILH 1002  
186 GLPVASARGREIILGPADNFEQGMRLAPITAYSCQTRGLGCIITSLTGRDKQVEGE 245  
1003 GLPVASARGREIILGPADNFEQGMRLAPITAYSCQTRGLGCIITSLTGRDKQVEGE 1062  
246 VQVAVSTATQSFLLATCVNGVCTVTFHAGSKTLAGPKPITQMTYNNVDOLVGCAPFGAR 305  
1063 IQVLSVYQTQFLGTSISGVMTVYHAGNKTLLAGSGPVTQMTYNSAEGDVLGWPSPGTX 1122  
306 SMTPTCGSSDLYLVTRHADIVPVRARGSGSLSPRPVSYLKSGSGGPLLCPSSGAVG 365  
1123 SLBECTGAVDYLVTNRADVI PARRRGRGALLPRPLSTLKSGSGGVLCPRSHAVG 1182  
366 IFRAAVCTRGVAKAVDFIPVSMETMR 393  
1183 VFRAAVCTRGVAKSIDIPVETLIDIVTR 1210

RESULT 11  
GNWYJ8

genome polyprotein - hepatitis C virus (strain HC-J8)  
N:contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C:Species: hepatitis C virus  
C:Accession: A40250; P00397; P00559  
C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 19-Jan-2001  
R:Okamoto, H.; Kural, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;  
Virology 188, 331-341, 1992  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to report  
A:Reference number: A40250; MUID:9223032; PMID:1314459  
A:Accession: A40250

A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609  
R:Chen, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.;  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to ex  
A:Reference number: P00393; MUID:92266871; PMID:1316939

A:Accession: P00397  
A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DBJ:D10134  
A:Experimental source: isolate E-b12  
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,  
Blochm. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: P00554; MUID:92068204; PMID:1720309

A:Accession: P00559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <XAT>  
A:Cross-references: GB:D10542; GB:D090518; NID:G221523; PIDN:BA01418.1; PID:G221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolyase; nonstructural  
F:111-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <CPC>  
F:116-191/Product: major envelope protein E #status predicted <MEE>  
F:330-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1667-2017/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,223,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,235

Query Match 68.2%; Score 1401; DB 1; Length 3033;  
Best Local Similarity 63.4%; Pred. No. 9e-107;  
Matches 246; Conservative 67; Mismatches 75; Indels 0; Gaps 0;

6 GGAVFGLALLTSPYKVLARLIMLQYLITRVEAHLOVWIPNLVNRGRDAIILITLC 65  
823 GAAVLVITLFTLPGYKTLISRFMLCYLLLAEMVQEMAPMVQVGRGDIIMAV 882  
66 AVHDELIFDITKLLAIFGLMLVQAGITVYFVRAQGLIRACMLVRKAAGHYVMAR 125  
883 ILHPLVFEVTKMLLALIGPAYLLKSLIRIPYVRAHALRMTVVRHLAGRYYQVNL 942  
126 MKLAALGTIVYDHLTPLODMAHAGLRDLAVAVEPVFSDMEVKIITWGDPTAACGDIIS 185  
943 IIRGRMTGTYIDHLTPMSMANGLRDLAVAVEPVFSDMEVKIITWGEETIACGDIILH 1002  
186 GLPVASARGREIILGPADNFEQGMRLAPITAYSCQTRGLGCIITSLTGRDKQVEGE 245  
1003 GLPVASARGREIILGPADNFEQGMRLAPITAYSCQTRGLGCIITSLTGRDKQVEGE 1062  
246 VQVAVSTATQSFLLATCVNGVCTVTFHAGSKTLAGPKPITQMTYNNVDOLVGCAPFGAR 305  
1063 IQVLSVYQTQFLGTSISGVMTVYHAGNKTLLAGSGPVTQMTYNSAEGDVLGWPSPGTX 1122  
306 SMTPTCGSSDLYLVTRHADIVPVRARGSGSLSPRPVSYLKSGSGGPLLCPSSGAVG 365

Db 1123 SLDPCTCGAVDLYLVTNRMDVIFPRKDRRGLISPRPLSTLKSSGGPVLCSRGHVG 1182  
 QY 366 IFRAAVCTRGVAKAVDFIPVESMETMR 393  
 Db 1183 LFRPAVCAAGVAKSIDIFIPESLDVATR 1210

## RESULT 12

T08841  
 polyprotein - douroucouli hepatitis GB virus A  
 C/Species: douroucouli hepatitis GB virus A  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C/Accession: T08841  
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08841  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-3005 <ERK>  
 A/Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAQ40501.1; PID:g2828600  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 19.9%; Score 408.5; DB 2; Length 3005;  
 Best Local Similarity 31.3%; Pred. No. 6.5e-25;  
 Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 54 RGGRD-----AIIILTCVHPELIFDITKLLAIFGLMVLQAGITKVPYFRAQGLIRAC 109  
 Db 813 RGGRDVTVAVWVAAIGLIFREVRCALTA-----LAALLDSIDVLETL-ILTA 864  
 QY 110 MLVKA-----AGHYVQMAFMKLAITGYVVDHLTLPQMAHAGLDAVAPEV 161  
 Db 865 QPAAARLLDSLTFGLADLTFARVRLERGVTLFPHCGQVSGAAAILXDLGVALLPEV 924  
 QY 162 IFSMEKILTWGDTAACGDIISGLVSAARGEILLG--PDNFGSGCRLLAPITAY 219  
 Db 925 SVTARDCVYVADARFALACGQVAGLGVVAKSGEVLVGVFSPRALPPGFVPAVYV- 993  
 QY 220 SQQTRGLIGITITLSTGKQVGEVQVSTATQSEFLATCVNGCVTFEGASKTLAG 279  
 Db 984 MQRLGFPVSVKTSMLGRDREHSGIVLGTSTTRSMGTCVNGVMYTTFHGSNARTLAG 1043  
 QY 280 PKGPTTQMTYVDDDLVGMQAPPRARSMPTCTGSSDPLVLTNRADVLPVRRGDSRSL 339  
 Db 1044 PVGVNCRWSPSDVAVYPLPSGASCIPECKCTQSVCCIRN--DALCHGRSLKVEL 1101  
 QY 340 LSPRPVSYLKSSGSGPLCPSGHVGIFRAVCTRGV-----AKAVDFIPVES 387  
 Db 1102 DLPRFISDFRSSSGSPILCDGHHVGMV-VSLHKGVMGVKVMETLPKDS 1155

## RESULT 13

T08839  
 polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: T08839  
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <ERK>  
 A/Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAQ40501.1; PID:g2828598  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 16.7%; Score 342.5; DB 2; Length 2970;

Best Local Similarity 28.9%; Pred. No. 1.8e-19;  
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;

QY 78 LLALFGP-----LWVLAQGITKVPYFRAQGLIRACMLVRAAGHYVQMAFVKLA 129  
 Db 816 LVAAFWFMRREAAVCAFIPLFGFEDVDVILEVLVSSPNVLARLADSLVAAGDKLA 875  
 QY 130 ALTGTYVDHLTLPD--MAAG-----LRDLAAVPEVIESDVEVITMGADTA 178  
 Db 876 T--TWLVEKRRKRCNCFVLAHAGOVTRTAQRLQWGMALBPVAVHPEDCAVMDARTL 932  
 QY 179 ACGDIIISGLPVSARREIILGPADNFBQGMRL-----LAPITAVSQQTRGLIGIIT 232  
 Db 933 SCGQSVHKKPVVARRGDEVLLGVNGV--WELPPGFVPAVYVH-HHGKGFPGVVK 987  
 QY 233 SLTGRDKQVGEVQVSTATQSEFLATCVNGCVTFEGASKTLAGKSPITQMTYVVD 292  
 Db 988 SMTGMDTEHVNQVNVLLGTSTTRSMGTCVNGVMYTTGHSNARTLAQMGVNSRWMSAS 1047  
 QY 293 ODVGMQAPPGARSMPTCTGSSDLYLVTNRADVLPVRRGDSRGLS----- 341  
 Db 1048 DVAVVYPLPVGAKCLEPKCKPQGVVVI-----RND--GALCHGTIGRTVELDL 1094  
 QY 342 PRPVSYLKSSGSGPLCPSGHVGIFRAVCTRG-----VAKAVDFIPVESMET 391  
 Db 1095 PALLCDPRGSSGSPILCDGHHVGMV-LSVLHRSRVTVGIRYKEMETLPREAITH 1150

## RESULT 14

H71426  
 hypothetical protein - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 A/Variety: Columbia  
 C/Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
 C/Accession: H71426  
 R/Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Poll, T.M.; Terry, N.; Giele  
 avenagh, T.; Hempel, S.; Kotter, P.; Eitlan, K.D.; Rieger, M.; Schaeffer, W.; Funk, B.  
 Nature 391, 485-488, 1998  
 A/Authors: Wellenreger, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdemonech,  
 etnot, A.; Montes, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansk  
 C.; Chalmers, N.  
 A/Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal:  
 A/Reference number: A71400; MUID:98121113; PMID:9461215  
 A/Accession: H71426  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-692 <REV>  
 A/Cross-references: GB:Z97340; NID:g2244950; PID:e327492; PID:g2244965  
 C/Genetic:  
 A/Map position: 4COP9-4G3845

Query Match 5.5%; Score 112; DB 2; Length 692;  
 Best Local Similarity 24.1%; Pred. No. 0.3;  
 Matches 97; Conservative 45; Mismatches 135; Indels 126; Gaps 23;

QY 26 LARLLWLOLYLTRVAHQVMPVLANVGRDAILLTCAVHPELIPITKLLAIFGP 85  
 Db 83 IADLAFGINVLMR-----QGNFTTSVAVAGSCEL-----KPELIMDLTEILR--FLT 131  
 QY 86 LNVLAQGITKVPY--FVRAQGLIRACMLVR--KAAAGHYVQMAFV----- 126  
 Db 132 LCMV---FSKRPVAVLEBSAGYHEDVLLQKPRAGVGHIMQRAFTIIRDTNSKILLR 188  
 QY 127 -----KLAALGTYYVDHLTLPQMAHAGLDAVAVPEVIESDVEVITMGADTA 179  
 Db 189 GTSHIKDTLTAAGAVAPFHSVLAHD--GGLSNLVLGY----- 226  
 QY 180 GPDIIISGLPVSARREIILGPADNFBQGM--RLAPITAVSQQTRGLIGIITSLGR 237  
 Db 227 CG-----MVAAAR-----WAKLSVP-----CLKKL----- 248  
 QY 238 DKNQVGEVQVSTATQSEFLATCV-----NGVCWTVFHGASKTLAGKGP--I 284

```
Db      249 DENP-SFNVQIVGSHSLGGTASLITYIREQKERASATCTFPAGTNPMLINSGSKHFI 307
Qy      285 TOMTNNVDQDLY---GWQAPPGASMTPTCTGSSDLYLVTRHADVIP-VRRGDSRGSL 340
Db      308 TTIING--SDLVPTFSASSVDLSEVTSSSWSNDLRDQVEHTRVLSVYRSATATIGSRL 365
Qy      341 SPRPVSYIKSSSGGELLCP--SGHVGIFRAAVCTRGVAKAYD 381
Db      366 PSIASAKAKVAGAGAILRPVSSGTQVAAPLVNGC--GKIKCID 406
```

## RESULT 15

VHMMH2

structural protein 2 precursor - hepatitis E virus (strain Burma)

C:Species: hepatitis E virus

C&gt;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1999

C/Accession: C40778

R/Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.

Virology 185, 120-131, 1991

A&gt;Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi

A:Reference number: A40778; MUID:92024067; PMID:1926770

A:Accession: C40778

A:Molecule type: genomic RNA

A:Residues: 1-660 &lt;TAM&gt;

A:Cross-references: GB:M73218; NID:g330023; PIDN:AAA45736.1; PID:g330026

A&gt;Note: the authors translated the codon CGC for residue 2 as Ala

C:Superfamily: hepatitis E virus structural protein 2

C:Keywords: structural protein

F:1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F:23-660/Product: structural protein 2 #status predicted &lt;SP2&gt;

## Query Match

Best Local Similarity 5.0%; Score 102.5; DB 1; Length 660;

Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

```
Qy      76 TKLL--AIFGLMVLQAG-----ITKPYFVR--AQLIPACMLVRKAAGHYQMA 124
Db      151 TNLVLAAPLSPPLPLQDGNTHIMATEASNYAQYVARATIRYRLVPNAVGVATISIS 210
Qy      125 FMKLAALGTIVVDHLTPLODWAHAGLDLAVAVEPIFSDMEKTIITWQADTAACDII 184
Db      211 FWPQTITPTTSV-----DNMSITDVRILVQPIASELYI----- 246
Qy      185 SGLPVASRRGRREILGPAD--NFEQGQRLAPI-TAYSQQTRGLL-----GCITSLTG 236
Db      247 -----PSERLHYRNQGMRSVETSGVAEEBATSGLVWLCTHGSIVNSYTN 290
Qy      237 -----RDXNQEVEGVVSTATQSFL 257
Db      291 TPYTGAIGLDFALELEFRNLTPGNTNTRVSRYSSTARHLRGRADGTALTTTAAATRFM 350
Qy      258 A---TCVNGV-----CWTVFH-----GAG-----SKTLGAPKG-PTT 285
Db      351 KDLYFTSTNGVGEIRGIALTLFNLADTLGLPTELISAGQLFYSRPVVSANGEPV 410
Qy      286 QMTYNNVDQDLYVQWAPPGASMTPTCTGSSDLYLV--TRHADVIPVRRGDSRG-SLIS 341
Db      411 KLTIVSVENA---QQDKGIALPHDIDGESRVVIQDYDQHEQDRPTSPAPSRPFSVL 466
Qy      342 PRPVSYLK-----GSSGGLPLCPSGHVGIFRAAVCTRGVAKAYDFIPV 385
Db      467 ANDVLWLSLVAARYDQSTYGSSTGCPYV--SDSVTLVNVATGAQAVARSLDWTKV 519
```

Search completed: May 6, 2004, 09:37:17  
Job time: 14.7992 secs

Fri May 7 13:37:01 2004

us-10-650-585-11.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 8.20459 Seconds

(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-11

Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALITSP.....RGVAKAVDTPEVSEMTNR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1967	95.8	3010	P26662 h genome po
2	1950	95.0	3010	Q00289 h genome po
3	1935	94.3	3010	P29846 h genome po
4	1888	92.0	3010	P26663 h genome po
5	1764	85.9	3011	P26664 h genome po
6	1752	85.3	3011	P27958 h genome po
7	1403	68.3	3033	P26660 h genome po
8	1401	68.2	3033	P26661 h genome po
9	102.5	5.0	660	P27426 hepatitis e
10	102.5	5.0	660	P27426 hepatitis e
11	101.5	4.9	1780	P27426 hepatitis e
12	101.5	4.9	564	P27426 hepatitis e
13	101	4.9	600	P27426 hepatitis e
14	94	4.6	444	P27426 hepatitis e
15	93.5	4.6	1380	P27426 hepatitis e
16	93	4.5	434	P27426 hepatitis e
17	92.5	4.5	706	P27426 hepatitis e
18	92	4.5	444	P27426 hepatitis e
19	92	4.5	659	P27426 hepatitis e
20	92	4.5	3414	P27426 hepatitis e
21	91.5	4.5	401	P27426 hepatitis e
22	91.5	4.5	485	P27426 hepatitis e
23	91.5	4.5	660	P27426 hepatitis e
24	91.5	4.5	3412	P27426 hepatitis e
25	90.5	4.4	403	P27426 hepatitis e
26	90.5	4.4	3414	P27426 hepatitis e
27	88.5	4.4	961	P27426 hepatitis e
28	88.5	4.3	355	P27426 hepatitis e
29	88.5	4.3	3432	P27426 hepatitis e
30	88	4.3	441	P27426 hepatitis e
31	87.5	4.3	3432	P27426 hepatitis e
32	87.5	4.3	3432	P27426 hepatitis e
33	86.5	4.2	347	P27426 hepatitis e

34	86.5	4.2	378	1	NRAM_IATRA	P05803 influenza a
35	86	4.2	438	1	GALE_MEICO	O05026 neisseria g
36	86	4.2	433	1	DCUA_MOLSU	O34245 wolfeella s
37	86	4.2	790	1	RIRI_HSVB	P28846 equine herpes
38	85.5	4.2	355	1	VP71_MYCTU	O50650 mycobacteri
39	85	4.1	470	1	NRAM_IATRA	P03472 influenza a
40	85	4.1	730	1	HELS_METMA	O9p2r7 methanosarc
41	85	4.1	854	1	PMP2_SCHPO	O9c1x1 schizosacch
42	85	4.1	313	1	CHRS_RAT	O88278 rattus norv
43	84.5	4.1	453	1	NRAC_BACCU	O07553 bacillus su
44	84.5	4.1	1705	1	PIPV_MOUSE	P70289 mus musculu
45	84	4.1	305	1	OPFC_BACCU	P24139 bacillus su

#### ALIGNMENTS

RESULT 1  
POLG\_HCVUA STANDARD; PRT; 3010 AA.  
ID POLG\_HCVUA  
AC P26662;  
DT 01-AUG-1992 (rel. 23, Created)  
DT 01-AUG-1992 (rel. 23, Last sequence update)  
DT 28-FEB-2003 (rel. 41, Last annotation update)  
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Japanese) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitis C virus.  
NCBI\_TaxID=11116;  
OX  
RN  
RP  
RX MEDLINE=91088550; PubMed=2175903;  
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
RA Sugimura T., Shimotohno K.;  
RT "Molecular cloning of the human hepatitis C virus genome from  
RT Japanese patients with non-A, non-B hepatitis".  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528 (1990).  
RN [2]  
RP  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,  
RA Ohkoshi S., Shimotohno K.;  
RT "Molecular structure of the Japanese hepatitis C viral genome.";  
RL FBS Lett. 280:325-328 (1991).  
CC  
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA) (N).  
CC  
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein C and RNA.  
CC  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D90208; BAA14233.1; -

DB 1; Length 3010;

-----

---

FT	CARBOHYD	5/6	5/6	N-LINKER
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	86
87	87	87	87</	

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

FT	CARBOHYD	233	233	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	234	234	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	250	250	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	305	305	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	417	417	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	423	423	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	430	430	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	448	448	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	532	532	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	540	540	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	556	556	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	576	576	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	623	623	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	645	645	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	2041	2041	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	2077	2077	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	2240	2240	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	2529	2529	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	2788	2788	N-LINKED <th>(GLCNAC .)</th> <th>(POTENTIAL)</th>	(GLCNAC .)	(POTENTIAL)
SQ	SEQUENCE	3010 AA;	327047 MW;	AAD26795CDFE215 CRC64;		
Query Match 94.3%; Score 1935; DB 1; Length 3010;						
Best Local Similarity 91.9%; Pred. No.7e-149;						
Matches 361; Conservative 17; Mismatches 15; Indels 0; Gaps 0						
QY	1	MAASCGGAVF	IGLALITLSPYKYLARLWMLOYLITRVEAHLOVMIPLNVRGGRDAI	60		
DB	814	MAASCGGAVF	GLVLTLSPHYKMFARLIMWLQYPTTRBAHLQWVIPPUNRGGRDAI	873		
QY	61	ILITCAVHBE	LFDITTKLLALFQPLWLOQIGIKVYFFRAGLIRACMLVRKAAGAY	120		
DB	874	ILITCAVBE	LLFDITTKLLALFGLWLOGLRIYFPRAGLIRACMLVRKAAGAY	933		
QY	121	VQMAFMKLA	LALGTIVYDHLITPLQDMAHAGLRDLAVAEVIRFSDMEVKIITWGAATAAC	180		
DB	934	VQMAFMKLA	LALGLTVYDHLITPLQDMAHAGLRDLAVAEVIRFSDMEVKIITWGAATAAC	993		
QY	181	GDIISGLP	VSANRGEIILGPADNEEGQWRLLAPITAYSQOTRGLIGCTITSITGRDN	240		
DB	994	GDIISGLP	VSANRGEIILGPADNEEGQWRLLAPITAYSQOTRGLIGCTITSITGRDN	1052		
QY	241	QVEGEVQV	STATQSEFLATCNQVCMVYFHGAGSKTLAGKSPITQMTYNNVDDLVGMA	300		
DB	1054	QVEGEVQV	STATQSEFLATCNQVCMVYFHGAGSKTLAGKSPITQMTYNNVDDLVGMA	1112		
QY	301	PGARSMT	PCTGSSDLYLTRADVIPRRRGDSRGLSPRPVSYLKSGSGGPIICPS	360		
DB	1114	PGARSMT	PCTGSSDLYLTRADVIPRRRGDSRGLSPRPVSYLKSGSGGPIICPS	1172		
QY	361	GHAVGIP	RAAVCTRVAKADPFIYESMETMTR	393		
DB	1174	GHAVGIP	RAAVCTRVAKADPFIYESMETMTR	1206		
RESULT 4						
POLG_HCVXK STANDARD; PRT; 3010 AA.						
AC	P26663;					
DT	01-AUG-1992	(Rel. 23, Created)				
DT	01-AUG-1992	(Rel. 23, Last annotation update)				
DT	10-OCT-2003	(Rel. 42, Last annotation update)				
DE	Genome polypeptide [contains: Capsid protein C (Core protein) (p22);					
DE	Envelope glycoprotein [contains: Glycoprotein E2					
DE	(GP68) (GP70) (NS1); Protein F7; Nonstructural protein NS2 (p21)					
DE	(EC 3.4.22.-); Protease/helicase NS3 (p70) (hepaticin)					
DE	(EC 3.4.21.98); Nonstructural protein NS4A (p4); Nonstructural protein					
DE	NS4B (p27); Nonstructural protein NS5A (p56); Nonstructural protein					
DE	NS5B (p66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].					
OS	Hepatitis C virus [isolate BX] (HCV).					
OC	Virusess; ssRNA positive-strand virusess, no DNA stage; Flaviviridae;					
OC	Hepadnavirus.					
XX	NCBI_Taxid=11105;					
NN	[1]					





```

RA MEDLINE=9117826; PubMed=1848704.
RX Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic Organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC CC -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the Pe
CC position. Cys or Thr in Pl and Ser or Ala in Pl'.
CC CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {N} (N).
CC CC -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M62321; AAA45676.1; -.
DR PIR; A39166; GNWVC3.
DR PDB; 1AIY; 16-FEB-98.
DR PDB; 1HEI; 25-NOV-98.
DR MEROPS; S29..001; -.
DR DR InterPro; IPRO09003; Cys_Ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4A.
DR InterPro; IPR001490; HCV_NS4B.
DR InterPro; IPR002868; HCV_NS5A.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; peptidase_C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PsYlr.
DR Pfam; PF01543; HCV_capsid_1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C_1.
DR Pfam; PF00998; Viral_RdRp_1.
DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDC; 1.
KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Envelope protein; Helicase; Arp-binding;
KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KM 3D-structure.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
CAPSID PROTEIN C (POTENTIAL).
MATRIX PROTEIN (POTENTIAL).
MAJOR ENVELOPE PROTEIN E (POTENTIAL).
NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCBAF9 CRC64;

Query Match 85.9%; Score 1764; DB 1; Length 3011;  
 Best Local Similarity 81.9%; Pred. No. 5.6e-135;  
 Matches 322; Conservative 34; Mismatches 37; Indels 0; Gaps 0;

1 MAASCGAVFGLALLSPYKYLARLWLOYLTRVHAHQVPIPLNRGGDAI 60  
 814 VAASCGGVAVGLALTLSPYKRYISWCLMWLOYLTRVHAHQVPIPLNRGGDAV 873  
 61 ILATCAHPELIPDITKLALIFGPLNVLQAGITKVFYFAAGLIRACMLVKAAGHY 120  
 874 ILLMCANHPPLVPIITKLALAVRPWLQASLIXVYFPRVQGLRFLCAARKMGHY 933  
 121 VQNAFMKLAITGYVVDHLPLDQMAHAGIRDLAVNEVYISDMEVKIIITGADTAAC 180  
 934 VQVWIKLGLALTYVYNHLPFLDMAHANGIRDLAVNEVYISDMEVKIIITGADTAAC 993  
 181 GDIISGLPVASARETEILGPADNFGQWRLLAPITAYVQOQRGLIGCITITLTGRDN 240  
 994 GDIISGLPVASARETEILGPADNFGQWRLLAPITAYVQOQRGLIGCITITLTGRDN 1053  
 241 QVEGEVAVSTATOSPLATCVNGVCMVTFHAGASKTLAGPKPIITQNTNVDDLVGMOA 300  
 1054 QVEGEVAVSTATOSPLATCVNGVCMVTFHAGASKTLAGPKPIITQNTNVDDLVGMOA 1113  
 301 PPGASMTPTCTGSSDLYIVTRHADVPVRRGDSRSLSISPRVSVLTKSSGGPILCP 360  
 1114 PPGASMTPTCTGSSDLYIVTRHADVPVRRGDSRSLSISPRVSVLTKSSGGPILCP 1173  
 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTM 393  
 1174 GHAVGIFRAAVCTRGVAKAVDFIVESMETTM 1206

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 DE Hepatitis C virus (isolate H) (HCV).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirinae.  
 OC NCBI\_Taxid=1108;  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauste G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M., Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C virus; comparison with American and Japanese isolates.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RA MEDLINE=9731322; PubMed=9187654;  
 RX Yao N., Heeson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RP MEDLINE=98154321; PubMed=9493270;  
 RX Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.  
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE. ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the p6 position. Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1 AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND NS5A.  
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M67463; AAA45534.1; -.  
 CC PIR: A36814; GNVVCH.  
 CC PDB: 1HRI; 25-NOV-98.  
 CC PDB: 1AIV; 16-FEB-99.  
 CC PDB: 1AIR; 17-JUN-98.  
 CC MEROPS: S29.001; -.  
 CC TRANSFAC: T04155; -.  
 CC InterPro: IPR009003; Cys\_Set\_trypsin.  
 CC InterPro: IPR001410; DED.  
 CC InterPro: IPR002522; HCV\_capsid.

Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length	Indels	Gaps
Query Match	85.3%	320	36	1752	1	3011	0	0
Best Local Similarity	81.4%	320	36	1752	1	3011	0	0
Matches	320	320	36	1752	1	3011	0	0
Conservative	36	36	36	1752	1	3011	0	0
Score	1752	1752	1752	1752	1	3011	0	0
DB	1	1	1	1	1	3011	0	0
Length	3011	3011	3011	3011	1	3011	0	0
Indels	0	0	0	0	1	3011	0	0
Gaps	0	0	0	0	1	3011	0	0
Query Match	85.3%	320	36	1752	1	3011	0	0
Best Local Similarity	81.4%	320	36	1752	1	3011	0	0
Matches	320	320	36	1752	1	3011	0	0
Conservative	36	36	36	1752	1	3011	0	0
Score	1752	1752	1752	1752	1	3011	0	0
DB	1	1	1	1	1	3011	0	0
Length	3011	3011	3011	3011	1	3011	0	0
Indels	0	0	0	0	1	3011	0	0
Gaps	0	0	0	0	1	3011	0	0

```

DB 1114 POGSRSLPTCTGSSDLYLVTRHADVTPVRRRGSRSLSPRISLYKSSGGLPCT 1173
QY 361 GHAVGIFRAVCTRGVAKAVDFIPVSEMERMTMR 393
DB 1174 GHAVGLFRAVCTRGVAKAVDFIPVSEMERMTMR 1206

RESULT 7
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
ID POLG_HCVJ6
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Genome glycoprotein [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepactivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11113;
RN
RP SEQUENCE FROM N. A.
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704 (1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position. Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D00944; BA00792.1; -
DR PIR; J01303; J01303.
DR HSSP; P27958; IHEI.
DR MEROPS; S29.001; -
DR MEROPS; U39.001; -
DR InterPro; IPR0030003; Cys_ser_lyspsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4A.
DR InterPro; IPR001490; HCV_NS4B.
DR InterPro; IPR002868; HCV_NS5A.
DR InterPro; IPR002166; HCV_RdRP.

```

```

DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4A; 1.
DR Pfam; PF01001; HCV_NS4B; 1.
DR Pfam; PF01506; HCV_NS5A; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
FT INIT_MER 1
FT CHAIN 1
FT CHAIN 115
FT CHAIN 116
FT CHAIN 191
FT CHAIN 192
FT CHAIN 383
FT CHAIN 384
FT CHAIN 734
FT CHAIN 1010
FT CHAIN 1011
FT CHAIN 1619
FT CHAIN 1620
FT CHAIN 1866
FT CHAIN 2017
FT CHAIN 2018
FT CHAIN 3033
FT CHAIN 347
FT CHAIN 369
FT ACT_SITE 1087
FT ACT_SITE 1087
FT ACT_SITE 1111
FT ACT_SITE 1111
FT ACT_SITE 1169
FT ACT_SITE 1230
FT NE_BIND 1234
FT SITE 1320
FT SITE 1323
FT CARBOHYD 196
FT CARBOHYD 209
FT CARBOHYD 234
FT CARBOHYD 234
FT CARBOHYD 305
FT CARBOHYD 417
FT CARBOHYD 423
FT CARBOHYD 430
FT CARBOHYD 448
FT CARBOHYD 477
FT CARBOHYD 534
FT CARBOHYD 542
FT CARBOHYD 558
FT CARBOHYD 578
FT CARBOHYD 627
FT CARBOHYD 649
FT CARBOHYD 1091
FT CARBOHYD 1091
FT CARBOHYD 2038
FT CARBOHYD 2811
FT CARBOHYD 2811
SQ SEQUENCE 3033 AA; 329165 MW; F957F5CIA273B93E CRC64;

```

```

Query Match 68.3%; Score 1403; DB 1; Length 3033;
Best Local Similarity 64.9%; Pred. No. 1.3e-105;
Matches 252; Conservative 58; Mismatches 78; Indels 0; Gaps 0;

```

```

QY 6 GCAVFGALTLTSPYKVLRLIMWLOYLTRVEAHQVAVIPPLNRRGGDAIILLTC 65
DB 823 GALLVLTILFTLTPGYKTLISRFIMWCYLLTLAEAMQEPAPMQVGGSDGITIMAYV 882
QY 66 AVHPELIPDITKLLAIFGLMVLQAGITKVEYFRAOGLIRACMLVRLAKAGHYVQAF 125
DB 883 IFCPGVPIITKMLLAVGPAVLKGLATRVYFVRAHLLMCMVMVRLHLAGRYVQWL 942
QY 126 KTLAALTGYVYDHLTPLODMAHAGRLDAVAVEPIYSDMEVKIITGCATLACGDIIS 185
DB 943 LALGRWTGYIVDHLTPMSDMANGRLDAVAVEPIIFSPMEKVIWGAETPAAGDILH 1002

```

QY 186 GLPVASARGSEILLGPADNFEQGMRLAPITAYSCOTRLGICITSLTGPRXNOVEGE 245  
 DB 1003 GLPVASARGSEILLGPADNFEQGMRLAPITAYSCOTRLGICITSLTGPRXNOVEGE 1062  
 QY 246 VOVASTGTGSEFLATGCVAGCMTVPHGAGSKTLGPKGPIQWTVNTDQDVLVGQAPEGAR 305  
 DB 1063 IGVASTGTGSEFLATGCVAGCMTVPHGAGSKTLGPKGPIQWTVNTDQDVLVGQAPEGAR 1122  
 QY 306 SMTPTCTGSSDLYLVTHADAVIPVRRGRSGSLSPRPVYKSGSGGELLCPGSHAVG 365  
 DB 1123 SLEBCTGAVDLYLVTHADAVIPVRRGRSGSLSPRPVYKSGSGGELLCPGSHAVG 1182  
 QY 366 IFPAVCTRGVAKAVDPVPEVSMETMR 393  
 DB 1183 VFPAVCTRGVAKAVDPVPEVSMETMR 1210

RESULT 8  
 POLG\_HCVJ8 STANDARD; PRT; 3033 AA.

AC P26661; 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA, positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 CX NCBI\_TaxID=11115;  
 OK [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9223023; PubMed=1314459;  
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Iizuka H., Tanaka T.,  
 RA Fukuda S., Tenda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes.";  
 RT Virology 188:331-341(1992).  
 RL CC  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D10988; BAA01761.1; -;  
 DR PIR: A40250; GNMVJ8.  
 DR HSSP: P27858; IHET.  
 DR MEROPS: S29.001; -;  
 DR MEROPS: U39.001; -;  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RBP.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVLtr.  
 DR Pfam: PF01543; HCV\_capsid\_1.  
 DR Pfam: PF01542; HCV\_core\_1.  
 DR Pfam: PF01539; HCV\_env\_1.  
 DR Pfam: PF01560; HCV\_NS1\_1.  
 DR Pfam: PF01538; HCV\_NS2\_1.  
 DR Pfam: PF02907; HCV\_NS3\_1.  
 DR Pfam: PF01006; HCV\_NS4a\_1.  
 DR Pfam: PF01001; HCV\_NS4b\_1.  
 DR Pfam: PF01506; HCV\_NS5a\_1.  
 DR Pfam: PF00998; Viral\_RBP\_1.  
 DR Pfam: PF08662; HCV\_NS1\_1.  
 DR SMART: SM00487; DEXDC\_1.  
 DR SMART: SM00487; DEXDC\_1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural  
 DR INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 2017  
 FT CHAIN 2018 3033  
 FT CHAIN 3034 369  
 FT TRANSMEM 347 1087  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT ACT\_SITE 1234 1241  
 FT NP\_SIND 1320 1323  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 FT CARBOHYD 3033 3033  
 FT SEQUENCE 330177 MW; 1A1Y3E7E3381FDIA CR664;  
 Query Match 68.2%; Score 1401; DB 1; Length 3033;  
 Best Local Similarity 63.4%; Pred. No. 1.9e-105;  
 Matches 246; Conservative 67; Mismatches 75; Indels 0; Gaps 0;

QY 6 GGVATGATLTSLPYRYVLLARLIMQLQYITRVEALHQLWIPPLNVRGRDAIILITC 65  
 DB 823 GLAIVTISITFLTPAYKILSRGVMISYVLVAEADIQQVPLLEVRGGRDGIWAV 882

QY 66 ANHEPILFDITKLLALFPGFLMVLQAGITVYFVFAQGLIRACMKVRKAGGHVYQMA 125  
 DB 883 ILHRLVFEVTKMLLALGPAYLLKASLRIPIFVRHALLRVCVLKHLIAGARYIQMLL 942  
 QY 126 MKLALGTGYVDLTPLODMAHAGLDLAVAVEPVIFSDMEVKIITWGDATACGDIIS 185  
 DB 943 IITIRMTGYIYDHLSPSLSTWMAQGLDLIAVEPVFSPMEKKVITWGETYACGDILH 1002  
 QY 186 GLPVASARGREILIGPDNEGCGKLLAPITAYSCQTRRLGCTITSLTGPRKQVEGE 245  
 DB 1003 GLPVASARGREILIGPDADGTYSKGMKLLAPITAYTCQTRGLGAIIVSLTGRKKNQOAG 1062  
 QY 246 VQVASTTQSGFLACVAGVCMVTFHAGSKTLGPKPITOMYTNVODDVLWQAPGAR 305  
 DB 1063 VQVASTTQSGFLACVAGVCMVTFHAGSKTLGPKPITOMYTNVODDVLWQAPGAR 1122  
 QY 306 SMPPTCGSSDLYLVRHADVIPVRRRGDSRGSILSPRPVSYLKSGSGPILCPSGHAYG 365  
 DB 1123 SLDPCTCGAVDLYLVTRNADYIPVRKDDRRGALLSPRLSTLKSGSGPILCPSGHAYG 1182  
 QY 366 IFRPAVCTRGVAKAVDPIPVESMETTR 393  
 DB 1183 IFRPAVCTRGVAKAVDPIPVESMETTR 1210

## RESULT 9

VST2\_HEVBU

ID\_VST2\_HEVBU STANDARD; PRT; 660 AA.

AC P29326;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Burma) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 OC Hepatitis E-like viruses.  
 OX NCBI\_TaxID=31767;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=92024067; PubMed=1926770;  
 RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,  
 RA Fry K.E., Reyes G.R.;  
 RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the  
 full-length viral genome."  
 RL Virology 185:120-131(1991).  
 CC -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M73218; AAA45736.1;  
 DR PIR: C40778; VHWMH2.  
 DR InterPro: IPR004261; SP2.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF03014; SP2; 1.  
 KM Signal.  
 FT SIGNAL  
 FT CHAIN  
 SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCCA461C CRC64;

Query Match 5.0%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 1.1; Indels 151; Gaps 19;  
 Matches 82; Conservative 52; Mismatches 130;

QY 76 TKLL--AIFGFLMVLQAG-----ITKVPYFVR--AAGLRACMLVRKAGGHVYQMA 124

DB 151 TLVLVYAAPSLPLDLOGTNTHMATASNVAQVRVARTIRPVLVNAVGVAAISIS 210  
 QY 125 FKLALIGTYVDLTPLODMAHAGLDLAVAVEPVIFSDMEVKIITWGDATACGDI 184  
 DB 211 FMPQTTTPTVS-----DNMSITSDVRLVQPGIASELVI----- 246  
 QY 185 SGLPVASARGREILIGPAD--NFEQGWRLAPI-TAYSCQTRGL-----GCITSLTG 236  
 DB 247 -----PSERLHYRQGMRSVETSGVAEEATSLGVMCIHSLVNSTN 290  
 QY 237 -----RDRNQVEGVQVSTQSTL 257  
 DB 291 TPYTGALGLDLELLEFRNLTPGNTNTRVRSSTASHRRRGADGAELETTATATRM 350  
 QY 258 A-----TCVNGV-----CWTVEH-----GAG-----SKTLAGPKG-PIT 285  
 DB 351 KDLYFSTNGVEIIGRLALTLFNLADTLGLPTELLISSAGGLFYRPPVVSANGEPTV 410  
 QY 286 QMTYNVQDVLVGMQAPPGARSMTPTCGSSDLYLV---TRHADVIPVRRRGDSRG-SLIS 341  
 DB 411 KLYTSVENA-----QDDKGLAIPHIDIDLGESRVYIDYDNQEHODRPTSPAPSRPFVLR 466  
 QY 342 PRPVSTLK-----GSGGFLCPSGHAYGIFRAAVCTRGVAKAVDPIPV 385  
 DB 467 ANDVYMLSLTAAEYDQSTYSGTGPVYV--SDSVTLVNVATGAQAVANSIDMTKV 519

## RESULT 10

VST2\_HEVPA

ID\_VST2\_HEVPA STANDARD; PRT; 660 AA.

AC P33426;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Pakistan) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 OC Hepatitis E-like viruses.  
 OX NCBI\_TaxID=33774;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=92115700; PubMed=1731327;  
 RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,  
 RA Malik I.A., Iqbal M., Purcell R.H.;  
 RT "Characterization of a prototype strain of hepatitis E virus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M80581; AAA45727.1;  
 DR InterPro: IPR004261; SP2.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF03014; SP2; 1.  
 KM Signal.  
 FT SIGNAL  
 FT CHAIN  
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;

Query Match 5.0%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 1.1; Indels 151; Gaps 20;  
 Matches 82; Conservative 54; Mismatches 128;

QY 76 TKLL--AIFGFLMVLQAG-----ITKVPYFVR--AAGLRACMLVRKAGGHVYQMA 124

```

Db 151 TNLVLAAPSLPLDQDGNTHMATEASVNAQYRAATIRPPLVPAVAGYALIS 210
Cc 125 FMKLAALGYVYVHLLPLQDMAGRLDLAVNEPIFGDMEXKITWAGDTAACDDI 184
Cc 211 FWPOTTTPTTSV-----DMNSITSDVRIWQFQIASELV----- 246
Cc 185 SGLVVSARRGREILLGPAD--NFGGCGRL-----APITAYSQ 221
Cc 247 -----PSEKHYNQKRSVETSGVAEEATSGVLWLCHISFPVSYTN 290
Cc 222 QT-RGLIGCI-----ITSITGRDKNQ-----VEGEVQVSTATQSL 257
Cc 291 TPYTGALGLDFALELEFEENLTPEGNTTRVRSVSTARHRLRGADGTALTTAATRFM 350
Cc 258 A-----TCVNGV-----CMTVEH-----GAG-----SKTLAPKXG-PIT 285
Cc 351 KDLFTSTNGVGEIRGIALTLFPLADTLGLFTLEISSAGGLFYRSRVVANGEPV 410
Cc 286 QMTNVNDODLVGWAQPPGASMPCTCGSSDLYLV--TRHADVTPVRRGDSRG-SLGS 341
Cc 411 KLYTSVENA-----QDDKGIAPHDIDGESRVVIGVDNQHEDQRPFPSPAPRPSVLAR 466
Cc 342 PRPYSYK-----GSSGCPILCPSGHANGITRAVCTRGVAKAVDFIPV 385
Cc 467 ANDVLMWLSLTAAYDQSTGYSSSTGPVYV--SDSVTLNVNATGAQAVARSLDMTKV 519

RESULT 11
POLG_MVEV STANDARD; PRT; 1780 AA.
AC P05769;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (core protein); Matrix
DE protein (envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2A, and NS2B; Protease/helicase (EC 3.4.21.98) (NS3)]
DE (Fragment).
OS Murray valley encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11079;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86200215; PubMed=3009829;
RA Dalgarno L., Trent D.W., Strauss J.R., Rice C.M.;
RT "Partial nucleotide sequence of the Murray Valley encephalitis virus
RT genome. Comparison of the encoded polypeptides with yellow fever
RT virus structural and non-structural proteins.";
RL J. Mol. Biol. 187:309-323(1986).
RN [2]
RP CARBOHYDRATES-LINKAGE SITES OF NS1, AND DISULFIDE BONDS.
RX MEDLINE=21405829; PubMed=11514736;
RA Blitvich B.J., Scanlon D., Shiell B.J., Mackenzie J.S., Pham K.,
RA Hall R.A.;
RT "Determination of the intramolecular disulfide bond arrangement and
RT biochemical identification of the glycosylation sites of the
RT nonstructural protein NS1 of Murray Valley encephalitis virus.";
RL J. Gen. Virol. 82:2251-2256(2001).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (see http://www.isb-sib.ch/announce/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc EMBL; X03467; CAA27184.1; -
Cc FIR; A24635; GNMVWV.
Cc HSSP; P14336; 1SVB.
Cc MEROPS; S07.001; -
Cc InterPro; IPR009003; Cys_Ser_trypsin.
Cc InterPro; IPR001122; Flavi_capsidC.
Cc InterPro; IPR000336; Flavi_glycoprote.
Cc InterPro; IPR000069; Flavi_M.
Cc InterPro; IPR001157; Flavi_NS1.
Cc InterPro; IPR000752; Flavi_NS2A.
Cc InterPro; IPR009487; Flavi_NS2B.
Cc InterPro; IPR002535; Flavi_Protep.
Cc InterPro; IPR007110; 1g-1like.
Cc InterPro; IPR001850; Peptidase_S7.
Cc Pfam; PF01003; Flavi_capsid_1.
Cc Pfam; PF02832; Flavi_glycop_C_1.
Cc Pfam; PF00869; Flavi_helicase_1.
Cc Pfam; PF00949; Flavi_helicase_1.
Cc Pfam; PF01004; Flavi_M_1.
Cc Pfam; PF00948; Flavi_NS1_1.
Cc Pfam; PF01005; Flavi_NS2A_1.
Cc Pfam; PF01002; Flavi_NS2B_1.
Cc Pfam; PF01570; Flavi_Protep_1.
Cc ProDom; PD001556; Flavi_glycoprote_1.
Cc ProDom; PD001496; Flavi_NS1_1.
Cc KX Polyprotein; Glycoprotein; Core protein; Coat protein;
Cc KM Envelope protein; Hydrolyase; Helicase; ATP-binding; Transmembrane;
Cc KW Nonstructural protein.
Cc INIT_MER 1
Cc CHAIN 1
Cc PROPEP 122 121
Cc CHAIN 218 292
Cc CHAIN 293 793
Cc CHAIN 794 1207
Cc CHAIN 1208 1372
Cc CHAIN 1373 1503
Cc CHAIN 1504 >1780
Cc TRANSMEM 44 60
Cc TRANSMEM 112 128
Cc TRANSMEM 278 294
Cc TRANSMEM 773 789
Cc TRANSMEM 1178 1194
Cc TRANSMEM 1219 1235
Cc TRANSMEM 1250 1266
Cc TRANSMEM 1312 1328
Cc TRANSMEM 1378 1394
Cc TRANSMEM 1401 1417
Cc TRANSMEM 1476 1492
Cc TRANSMEM 295 322
Cc DISULFID 352 408
Cc DISULFID 366 397
Cc DISULFID 384 413
Cc DISULFID 482 580
Cc DISULFID 597 628
Cc DISULFID 797 808
Cc DISULFID 848 936
Cc CARBOHYD 972 1016
Cc CARBOHYD 73 73
Cc CARBOHYD 140 140
Cc CARBOHYD 446 446
Cc CARBOHYD 923 923
Cc CARBOHYD 968 968
Cc CARBOHYD 1000 1000
Cc NON_TER 1780 1780
Cc SEQUENCE 1780 AA; 194866 MW; 0D6A47F0FB706DE CRC64;

```



```

RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Sconeing T., Pepin K., Spieth J., Sektion M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney B., Danne M.,
RA Dutt H., Edwards J., Fryman J., Hakensens B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozesky P., Riley A., Stromatt C.,
RA Wagner-McPherson C., Woliam A., Yoakum M., Bell M., Dedina N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kitchoff K., Toth K., King L., Bahret A., Miller B., Marx M.A.,
RA Martienssen R., McComble W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Voickert G., Wambolt R., Duesterhoef A., Stiekema W., Pohl T.,
RA Entian K.D., Terry N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCalligh B., Robben J., Grymopre B., Zimmermann W.,
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Stevenen M., Dirkse W., Moolman P., Klein Lankhorst R.,
RA Feldensieger T., Boche G., Rose M., Hauf J., Bernieser S., Hempel S.,
RA Weitzpusch M., Lamberth S., Villarroel R., Gielen J., Adliss W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K., Radd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.,
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RA thaliana."
RL Nature 408:823-826(2000).
CC -1- FUNCTION: May target chloroplast proteins to either the thylakoid
CC or envelope membranes.
CC -1- SUBCELLULAR LOCATION: Chloroplast stroma.
CC -1- TISSUE SPECIFICITY: Most abundant in green shoot tissue and
CC lower levels seen in the roots and etiolated buds.
CC -1- SIMILARITY: Belongs to the GTP-binding SRP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z21970; CAA79981.1; -
DR EMBL; AF092168; AAC64139.1; -
DR EMBL; AL162873; CAB85514.1; -
DR PIR; S36637; S36637.
DR HSSP; C07347; 1FPH.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54.SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54_1.
DR Pfam; PF02881; SRP54_N_1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SMO0382; AAA; 1.
DR TIGRFAWS; TIGR00659; fth; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;
KW Transil peptide.
FT FT TRANSIT 1 75
FT CHAIN 76 564
FT SIGNAL RECOGNITION PARTICLE 54 kDa
FT PROTEIN.
FT DOMAIN 76 370
FT DOMAIN 371 564
FT NP_BIND 183 190
FT NP_BIND 265 269
FT NP_BIND 323 326
FT CONFLICT 76 76
FT SEQUENCE 564 AA; 61232 MW; 423PF1285F89063E4 CAC64;
Query Match 4.9%; Score 101; DB 1; Length 564;
Best Local Similarity 26.1%; Pred. 1.2;
Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11.
QY VHEPEI-----FDIKLLLAIGPLMTVLAQGI-----TKVPEYFVACGILPACIMVVR 113
DB 154 VHDIVLKLMSGEVSEVLELQFAKSGPIYVILLAGIQGVGKTVCAKLACVYLKKQG--KSCMI- 210

```

QY 114 KAAGHYQVAFMKLAL--TGYVDHLTLPQ--DMAAGRLDAVAVEPVI FSDMEV 168  
 DB 211 --ADYVRPAIDLVILGEGVPPYTAGTDVAPAIAGQLKEAK-----NNVD 261  
 QY 169 KIITWGADTAACGIIISGLPVSARGREIL----LCPDNFEGQKRLIAPTAVSQ 223  
 DB 262 VIM---DIAGRLIDKGMDELDVKKFNFPELVLVDMTQ--EAAALVTFNVEI 315  
 QY 224 RGLIGCIITSLTGRDKNOVEGEVVS 250  
 DB 316 -GITGALITKLDGDSRGALSLVKEVS 341

## RESULT 13

DPO2\_MOUSE STANDARD; PRT; 600 AA.

AC P33611;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).  
 GN POLA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RX SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.  
 RA MEDLINE=9216788; PubMed=8463324;  
 RA Hayazawa H., Izumi M., Tada S., Takada R., Masutani M., Ue M.,  
 RA Hanaoka F.,

RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA  
 RT polymerase alpha-primase complex and their gene expression during  
 RT cell proliferation and the cell cycle."  
 RL J. Biol. Chem. 268:8111-8122(1993).  
 CC -1- FUNCTION: May play an essential role at the early stage of  
 CC chromosomal DNA replication by coupling the polymerase  
 CC alpha-primase complex to the cellular replication machinery (By  
 CC similarity).

CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme  
 CC (subunits A, B, C and D), which is assembled throughout the cell  
 CC cycle. The largest subunit (subunit A) has DNA polymerase  
 CC activity, the two smallest subunits (subunits C and D) have DNA  
 CC primase activity. Subunit B binds to subunit A.  
 CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M  
 CC PHASE (BY SIMILARITY).

CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D13546; BAA02746.1; -  
 DR PIR; B46642; B46642;  
 DR WGI; WGI:99690; PolA2.  
 DR InterPro; IPR007200; DNA.pol.alpha.B.  
 DR Pfam; PF04058; DNA.pol.alpha.B.1.  
 KW DNA replication; Nuclear protein; Phosphorylation.  
 FT DOMAIN 101  
 FT DOMAIN 107  
 FT DOMAIN 115  
 FT SEQUENCE 600 AA; 66267 MW; 79F9ABEEF33FEBC CRC64;

Query Match 4.9%; Score 101; DB 1; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 1.3;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 105 LIRACMLVRKAAGHYQVAFMKLALT-----CTYYVDL-----TPIQDMA 147

DB 27 LAELCVLYRQTEGKAVSEILAFCTSGAKTCLTVIDLINFEYEVINKLSKAMSHASKDSG 86  
 QY 148 HAGRLDLVAVEPVI FSDMEVYIIWGAADTAACGDI--ISGLP-----VSARGREI 197  
 DB 87 FAGTRDI-VSIQELLAEAEETLSTYTSKGLKXVSTPEPLTKRSVAARSPQ-- 144  
 QY 198 LIGPADNFGQKRLIAPTAVSQTRGLICIIITSLGRDKNOVEGEVVSATQSGFL 257  
 DB 145 LISPSS-----FSPSATPSQK-----YTSRFRN-----GVYTFPGSAQ--- 178  
 QY 258 ATCVNAGVCMTHFGAGSKTL--AGPKGITQMYNTVDDIYG 297  
 DB 179 ----GLSWSGRGSGSVSLKVGDPPEPLTSYKAMFOQLMG 215

## RESULT 14

Y447\_XYLFA STANDARD; PRT; 444 AA.

AC Q9P8T1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical zinc metalloprotease Xf1047 (EC 3.4.24.-).  
 GN Xf1047.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 NC NCB1\_TaxID=23711;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=945C;  
 RX MEDLINE=20165717; PubMed=10910347;  
 RA Simpson A.V.G., Keimach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carrer D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frome M., Fullan L.R.,  
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohenstein J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Kriger J.E., Kuramae E.E., Laigret P., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.P., Matukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhami A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."  
 RL Nature 406:151-159(2000).  
 CC -1- COFACTOR: Zinc (probable).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (By similarity).  
 CC -1- SIMILARITY: Belongs to peptidase family M50B.  
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



Search completed: May 6, 2004, 09:31:50  
Job time : 9.20459 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 36.7566 Seconds  
(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MASCGAGAFIGIALTLTSP.....RGVAKAVDFIPVESMETTR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioplasmid:\*  
17: sp\_archaeoplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1986	96.7	3010	12 Q9J3F9	Q9J3F9 hepatitis C
2	1981	96.5	3010	12 Q9J3H7	Q9J3H7 hepatitis C
3	1979	96.4	3010	12 Q9J3H9	Q9J3H9 hepatitis C
4	1974	96.2	3010	12 Q9J3H6	Q9J3H6 hepatitis C
5	1973	96.1	3010	12 Q9J3H4	Q9J3H4 hepatitis C
6	1973	96.1	3010	12 Q9J3H3	Q9J3H3 hepatitis C
7	1971	96.0	3010	12 Q9J3H3	Q9J3H3 hepatitis C
8	1970	96.0	3010	12 Q9J3H3	Q9J3H3 hepatitis C
9	1968	95.9	1166	12 Q81755	Q81755 hepatitis C
10	1968	95.9	12284	12 Q81817	Q81817 hepatitis C
11	1968	95.9	3010	12 P89966	P89966 hepatitis C
12	1968	95.9	3010	12 Q9J3G6	Q9J3G6 hepatitis C
13	1967	95.8	3010	12 Q9J3A2	Q9J3A2 hepatitis C
14	1967	95.8	3010	12 Q9J3D7	Q9J3D7 hepatitis C
15	1967	95.8	3010	12 Q9J3X6	Q9J3X6 hepatitis C
16	1967	95.8	3010	12 Q9J3X5	Q9J3X5 hepatitis C

17	1966	95.8	3008	12 Q9J3F4	Q9J3F4 hepatitis C
18	1965	95.7	3010	12 Q9J3H0	Q9J3H0 hepatitis C
19	1964	95.7	3010	12 Q9J3H9	Q9J3H9 hepatitis C
20	1963	95.6	3010	12 Q9J3H9	Q9J3H9 hepatitis C
21	1963	95.6	3010	12 Q81760	Q81760 hepatitis C
22	1961	95.5	3010	12 Q9J3Y3	Q9J3Y3 hepatitis C
23	1961	95.5	3014	12 Q9J3Y3	Q9J3Y3 hepatitis C
24	1960	95.5	3010	12 P88803	P88803 hepatitis C
25	1960	95.5	3010	12 Q9J3Y0	Q9J3Y0 hepatitis C
26	1960	95.5	3010	12 Q9J3H5	Q9J3H5 hepatitis C
27	1959	95.4	3010	12 Q9J3X8	Q9J3X8 hepatitis C
28	1959	95.4	3010	12 Q9J3Y7	Q9J3Y7 hepatitis C
29	1958	95.4	3010	12 Q9J3Y5	Q9J3Y5 hepatitis C
30	1958	95.4	3010	12 Q9J3H6	Q9J3H6 hepatitis C
31	1958	95.3	3010	12 Q9J3H6	Q9J3H6 hepatitis C
32	1957	95.3	3010	12 Q9J3H6	Q9J3H6 hepatitis C
33	1955	95.2	3010	12 Q88826	Q88826 hepatitis C
34	1954	95.2	3015	12 Q9J3H5	Q9J3H5 hepatitis C
35	1952	95.1	3010	12 Q9J3Y4	Q9J3Y4 hepatitis C
36	1951	95.0	3010	12 Q9J3Y4	Q9J3Y4 hepatitis C
37	1951	95.0	3010	12 Q9J3Y4	Q9J3Y4 hepatitis C
38	1951	95.0	3011	12 Q9J3Y4	Q9J3Y4 hepatitis C
39	1950	95.0	3010	12 Q9J3Y6	Q9J3Y6 hepatitis C
40	1949	94.9	3010	12 Q9J3Y5	Q9J3Y5 hepatitis C
41	1947	94.8	3010	12 Q9J3Y5	Q9J3Y5 hepatitis C
42	1947	94.8	3010	12 Q9J3Y5	Q9J3Y5 hepatitis C
43	1946	94.8	1275	12 Q06642	Q06642 hepatitis C
44	1946	94.8	3010	12 Q9J3Y7	Q9J3Y7 hepatitis C
45	1946	94.8	3010	12 Q9J3Y8	Q9J3Y8 hepatitis C

## ALIGNMENTS

RESULT 1  
ID Q9J3F9 PRELIMINARY; PRT; 3010 AA.  
AC Q9J3F9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MD33;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease progression."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL: AF207774; AAF65964.1; --  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR PIR; P50329; P50329.  
DR HSSP; P27958; 1HE1.  
DR MEROPS; S29.001; --  
DR MEROPS; U39.001; --  
DR GO; GO:0016021; C: integral to membrane; IEA.  
DR GO; GO:0019028; C: viral capsid; IEA.  
DR GO; GO:0019031; C: viral envelope; IEA.  
DR GO; GO:0005524; F: ATP binding; IEA.  
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F: electron transporter activity; IEA.  
DR GO; GO:0003123; F: RNA binding; IEA.  
DR GO; GO:0003168; F: RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR00345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PsVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SMO0487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SO SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E60C7 C6C64;

Query Match 96.7%; Score 1986; DB 12; Length 3010;  
 Best Local Similarity 94.9%; Pred. No. 5,4e-157;  
 Matches 373; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

DR 1 MAASCGAVFGLALLTSPYKVLARLIMWLYITREALQVWIPPLNVRGGRDAI 60  
 814 MAASCGAVFGLALLTSPYKVLARLIMWLYITREALQVWIPPLNVRGGRDAI 873

DR 61 ILTCAVHPELIDITKLLAIFGLPMLVQAGITKVPYFRAQGLIRACMLVRKAGHY 120  
 874 ILTCAVHPELIDITKLLAIFGLPMLVQAGITRMPYFRAQGLIRACMLVRKAGHY 933

DR 121 VQMAFKMLALTGTYVDHITPIQDMAHAGRLAVAVPVI FSDMEVKIITMGADTAC 180  
 934 VQMAFKMLALTGTYVDHITPIQDMAHAGRLAVAVPVI FSDMEVKIITMGADTAC 993

DR 181 GDIISGLPVARSRRREILLGPADNPEGQGRLLAPITAAVSOQRLGCIITSLTRDGN 240  
 994 GDIISGLPVARSRRREILLGPADNPEGQGRLLAPITAAVSOQRLGCIITSLTRDGN 1053

DR 241 QVEGEVQVSTAFQSLFATVNGVCVTFVHAGSKTLAKGKPTITMYNNVQDLYWQA 300  
 1054 QVEGEVQVSTAFQSLFATVNGVCVTFVHAGSKTLAKGKPTITMYNNVQDLYWQA 1113

DR 301 PPGARSMPTCTGSSPLVLTTRADVIPIRRRDSGSLSPPVSYLKGSSGGPLTCS 360  
 1114 PPGARSMPTCTGSSPLVLTTRADVIPIRRRDSGSLSPPVSYLKGSSGGPLTCS 1173

DR 361 GHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
 1174 GHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 2  
 ID 0933H7 PRELIMINARY; PRT; 3010 AA.  
 AC 0933H7;  
 DT 01-OCT-2000 (TREMBLrel. 15; Created)  
 DT 01-OCT-2000 (TREMBLrel. 15; Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25; Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 NC NCBT\_Taxid=11103;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD15;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease progression."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AF207566; AAF65946.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: P00804; P00804.  
 DR PIR: P50329; P50329.  
 DR HSSP: P26663; IUXP.  
 DR GO:0016021; C:Integral to membrane; IEA.  
 DR GO:0019028; C:Viral capsid; IEA.  
 DR GO:0019031; C:Viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PsVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SMO0487; DEXDC; 1.



Db 1114 PPGARSLPCTCGSSDLYLVRHADVIPRRRGDSRLSPRVSYLKSGGFLPCPS 1173

QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETMR 393

Db 1174 GHVGVIFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 4

ID 09DTE6 PRELIMINARY; PRT; 3010 AA.

AC 09DTE6; 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

CC Hepacivirus.

CC NCBI\_taxid=11103;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=HCV1221;

CC Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,

CC Harabara T., Ohata Y., Kanai K., Maruo H., Baba K., Hijioka M.,

CC Mishihiro S.,

CC "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients

CC with hepatocellular carcinoma: the 'progression score' revisited.",

CC Submitted (SPP-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; AB049101; BAB18814.1; -.

CC PIR; A61196; A61196.

CC PIR; PQ0246; PQ0246.

CC PIR; PS0329; PS0329.

CC HSSP; P26663; IUXP.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006508; F:electron transport; IEA.

DR GO; GO:0006350; F:proteolysis and peptidolysis; IEA.

DR GO; GO:0019079; F:viral genome replication; IEA.

DR GO; GO:0019087; F:viral transformation; IEA.

DR InterPro: IPR009003; Cys\_Ser\_Typsin.

DR InterPro: IPR000345; CysC\_heme\_BS.

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV\_capsid.

DR InterPro: IPR002521; HCV\_core.

DR InterPro: IPR002513; HCV\_env.

DR InterPro: IPR002531; HCV\_NSI.

DR InterPro: IPR002518; HCV\_NS1.

DR InterPro: IPR000745; HCV\_NS4A.

DR InterPro: IPR001490; HCV\_NS4B.

DR InterPro: IPR002868; HCV\_NS5A.

DR InterPro: IPR002166; HCV\_RdRp.

DR InterPro: IPR001650; Helicase\_C.

DR InterPro: IPR004109; Peptidase\_C29.

DR InterPro: IPR007095; RNA\_pol\_DS\_PS.

DR InterPro: IPR007094; RNA\_pol\_PSVir.

CC Pfam; PF01543; HCV\_capsid.1.

CC Pfam; PF01542; HCV\_core.1.

CC Pfam; PF01539; HCV\_env.1.

DR Pfam; PF01560; HCV\_NSI.1.

DR Pfam; PF01538; HCV\_NS2.1.

DR Pfam; PF02907; HCV\_NS3.1.

DR Pfam; PF01006; HCV\_NS4a.1.

DR Pfam; PF01001; HCV\_NS4b.1.

DR Pfam; PF01506; HCV\_NS5a.1.

DR Pfam; PF00271; Helicase\_C.1.

DR Pfam; PF00298; Viral\_RdRp.1.

DR ProDom; PD186062; HCV\_NSI.1.

DR SMART; SM00487; DEXDC1.1.

DR PROSITE; PS00190; CYTOCHROME\_C.1.

KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;

KW Hydroxylase; Nonstructural protein; Polypeptide;

KW RNA-directed RNA polymerase; Transferase; Transmembrane.

SQ SEQUENCE 3010 AA; 327108 MW; DE182D810BF78EE4 CRC64;

Query Match 96.2% Score 1974; DB 12; Length 3010;

Best Local Similarity 94.9%; Pred. No. 5.4e-156;

Matches 373; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKYLARLIMLOYLITRVAHLQVIPPINAGRDAL 60

Db 814 MAASCGAVFGLVLTLSPPYKVFARLIMLOYLITRVAHLQVIPPINAGRDAL 873

QY 61 ILTCAVPELIFDTKLLAIFGPIMWLGITKVPYFAOGLIACMIVRAAGHY 120

Db 874 ILTCAVPELIFDTKLLAIFGPIMWLGITKVPYFAOGLIACMIVRAAGHY 933

QY 121 VQAMFMKLAALTGIVYDHLTPLODMAHAGRLDAVPEYIFSDMEVKIITWGADTAAC 180

Db 934 VQAMFMKLAALTGIVYDHLTPLODMAHAGRLDAVPEYIFSDMEVKIITWGADTAAC 993

QY 181 GDIIISGLFVSARRREILIGPADNPEGGRFLAPITAYSQTRGLICITSLTGDKN 240

Db 994 GDIIISGLFVSARRREILIGPADNPEGGRFLAPITAYSQTRGLICITSLTGDKN 1053

QY 241 QVEGEVQVSTATOSFLATCVNGVCMVTFHAGAGSKTAGPGKPTQWTVNDQDLVMOA 300

Db 1054 QVEGEVQVSTATOSFLATCVNGVCMVTFHAGAGSKTAGPGKPTQWTVNDQDLVMOA 1113

QY 301 PPGARSLPCTCGSSDLYLVRHADVIPRRRGDSRLSPRVSYLKSGGFLPCPS 360

Db 1114 PPGARSLPCTCGSSDLYLVRHADVIPRRRGDSRLSPRVSYLKSGGFLPCPS 1173

QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETMR 393

Db 1174 GHVGVIFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 5

ID 09DTE4 PRELIMINARY; PRT; 3010 AA.

AC 09DTE4; 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

CC Hepacivirus.

CC NCBI\_taxid=11103;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=HCV1150;

CC Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,

CC Harabara T., Ohata Y., Kanai K., Maruo H., Baba K., Hijioka M.,

CC Mishihiro S.,

CC "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients

CC with hepatocellular carcinoma: the 'progression score' revisited.",

CC Submitted (SPP-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF



PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL; AB049093; BAB18806.1; -.

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00804; P00804.

DR PIR; P03329; P03329.

DR HSSP; P26663; IUXP.

DR GO; GO:0016021; C: integral to membrane; IEA.

DR GO; GO:0019028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005524; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F: electron transporter activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F: serine-type peptidase activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0016740; F: transferase activity; IEA.

DR GO; GO:0006118; P: electron transport; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral transformation; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR000345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002745; HCV\_NS2.

DR InterPro; IPR000745; HCV\_NS4.

DR InterPro; IPR001490; HCV\_NS4B.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RdRp.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA pol\_D5\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4; 1.

DR Pfam; PF01001; HCV\_NS4B; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR ProDom; PD186062; HCV\_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR PROSITE; PS00190; CYTOCHROME C; 1.

KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.

QO POLYPROTEIN; 3010 AA; 327324 MW; 3DB6CF249BD1151C CR64;

Query Match 96.1%; Score 1973; DB 12; Length 3010;

Best Local Similarity 94.4%; Pred. No. 6,6e-156;

Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 1 MAASCGAVFTGLALTLSPYKYLRLIWLIOYLITRVAHLQWIPPLNVRGGDAI 60

Db 814 MAASCGAVFTGLALTLSPYKYLRLIWLIOYLITRVAHLQWIPPLNVRGGDAI 873

QY 61 ILTCAVHPELIPITKLLAIFGLNWLQAGIKVYFPAQGLIACQLVRAAGGHY 120

Db 874 ILTCAVHPELIPITKLLAIFGLNWLQAGIKVYFPAQGLIACQLVRAAGGHY 933

QY 121 VQMAFMKLAALGTYYVDHDLTFLQDMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAAC 180

Db 934 VQMAFMKLAALGTYYVDHDLTFLQDMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAAC 993

QY 181 GIIIGLPSVARSRRGELLGPADNFEQGWRLIARITAYSOQTRGLCCITSLTRPKN 240

Db 994 GIIIGLPSVARSRRGELLGPADNFEQGWRLIARITAYSOQTRGLCCITSLTRPKN 1053

QY 241 QVEGEVQVVSFATQSFATLVGVGCMFVHGAGSKTLGAPKPIQMTYNTNDQDLVGMQA 300

Db 1054 QVEGEVQVVSFATQSFATLVGVGCMFVHGAGSKTLGAPKPIQMTYNTNDQDLVGMQA 1113

QY 301 PGASMSPTCTGSSDLYLVTRHADVIVRRRRDGRSGSLSPRVSYLKSGSGGFLCP 360

Db 1114 PGASMSPTCTGSSDLYLVTRHADVIVRRRRDGRSGSLSPRVSYLKSGSGGFLCP 1173

QY 361 GHAVGIFRAVCTRGVAVKAVNDPIPVESMETMR 393

Db 1174 GHAVGIFRAVCTRGVAVKAVNDPIPVESMETMR 1206

RESULT 6

QO 0807P3 PRELIMINARY; PRT; 3010 AA.

AC 0807P3

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Polypeptide.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepatitis C virus.

ON NCBI\_Taxid=11103;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MLE.

RX MEDLINE-22047193; PubMed-12051758;

RA Kishine H., Sugiyama K., Hijioka M., Kato N., Takahashi H., Nochi T., Nio Y., Hosaka M., Miyamari Y., Shimotohno K.;

RT "Subgenomic replicon derived from a cell line infected with the hepatitis C virus."

RL Biochem. Biophys. Res. Commun. 293:993-999 (2002).

DR EMBL; AB080299; BAC54896.1; -.

DR GO; GO:0019028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005524; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F: electron transporter activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F: serine-type peptidase activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0006118; F: electron transport; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P: transcription; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral transformation; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR000345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR000745; HCV\_NS4.

DR InterPro; IPR001490; HCV\_NS4B.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RdRp.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA pol\_D5\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; Helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF01606; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HelicC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polyprotein.  
 SQ SEQUENCE 3010 AA; 327097 MW; EE6418C7A23E686 CRC64;

Query Match 96.1%; Score 1973; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 6.6e-156;  
 Matches 372; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIMWLOYLITREAHLOQWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFVGLALTLSPYKVLARLIMWLOYLITREAHLOQWIPPLNVRGGRDAI 873  
 QY 61 ILTCAVPELIPDITKLLAIFGLPMLVLAQITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 874 ILTCAVPELIPDITKLLAIFGLPMLVLAQITKVPYFRAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFMKLAALTGYVVDHLPLODMAHAGRLDAVAVEVPSDMVKIITWGAADTAAC 180  
 DB 934 VQMAFMKLAALTGYVVDHLPLODMAHAGRLDAVAVEVPSDMVKIITWGAADTAAC 993  
 QY 181 GDITSGLPVSARRREILIGPADNFEQGRRLAPITAVSQQTRGLGCIITSLTGRDN 240  
 DB 994 GDITSGLPVSARRREILIGPADNFEQGRRLAPITAVSQQTRGLGCIITSLTGRDN 1053  
 QY 241 QVGEVQVSTAFATQVNGVMTVEFGAGSKTLAGKPIITOMYNNVDOLVGMQA 300  
 DB 1054 QVGEVQVSTAFATQVNGVMTVEFGAGSKTLAGKPIITOMYNNVDOLVGMQA 1113  
 QY 301 PPGARSMTPCTCGSSDLYLTRADYIVARRGDSRGLSPRPVSYLKGSSGGLLCPSS 360  
 DB 1114 PPGARSMTPCTCGSSDLYLTRADYIVARRGDSRGLSPRPVSYLKGSSGGLLCPSS 1173  
 QY 361 GHAVGIFRAVCTRGVAKANDPFPVSMETMR 393  
 DB 1174 GHAVGIFRAVCTRGVAKANDPFPVSMETMR 1206

## RESULT 7

Q93JH3 PRELIMINARY; PRT; 3010 AA.

AC Q93JH3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MD19;  
 RA Nageyama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL; AF207760; AA65950.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; PS0329; PS0329.

DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; F:electron transport; IEA.  
 DR GO; GO:0006308; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019079; F:transcription; IEA.  
 DR GO; GO:0019087; F:viral genome replication; IEA.  
 DR GO; GO:0019003; Cys Ser. Cyspsn.  
 DR InterPro; IPR000345; CytC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; peptidase\_C29.  
 DR InterPro; IPR007095; RNA pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA pol\_PSVIR.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; Helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF01606; HCV NS1; 1.  
 DR ProDom; PD186062; DEXDC; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327234 MW; 44C34677649C88D CRC64;

Query Match 96.0%; Score 1971; DB 12; Length 3010;  
 Best Local Similarity 93.9%; Pred. No. 9.7e-156;  
 Matches 369; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLALTLSPYKVLARLIMWLOYLITREAHLOQWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFVGLALTLSPYKVLARLIMWLOYLITREAHLOQWIPPLNVRGGRDAI 873  
 QY 61 ILTCAVPELIPDITKLLAIFGLPMLVLAQITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 874 ILTCAVPELIPDITKLLAIFGLPMLVLAQITKVPYFRAQGLIRACMLVRKAGGHY 933  
 QY 121 VQMAFMKLAALTGYVVDHLPLODMAHAGRLDAVAVEVPSDMVKIITWGAADTAAC 180  
 DB 934 VQMAFMKLAALTGYVVDHLPLODMAHAGRLDAVAVEVPSDMVKIITWGAADTAAC 993  
 QY 181 GDITSGLPVSARRREILIGPADNFEQGRRLAPITAVSQQTRGLGCIITSLTGRDN 240  
 DB 994 GDITSGLPVSARRREILIGPADNFEQGRRLAPITAVSQQTRGLGCIITSLTGRDN 1053  
 QY 241 QVGEVQVSTAFATQVNGVMTVEFGAGSKTLAGKPIITOMYNNVDOLVGMQA 300

DB 1054 QVEGEVQVSTAQSFATCTVNGVCMVTHGAGAKTLAGKPGITQMTNTNVDDLVGMOS 1113  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVIVPRRRGDSRGSLLSPVSYLKSSGGPLLCPS 360  
 DB 1114 PGARSLTPTCTCGSSDLYLVTRHADVIVPRRRGDSRGSLLSPVSYLKSSGGPLLCPS 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 8  
 Q68788  
 ID Q68788 PRELIMINARY; PRT; 3010 AA.  
 AC Q68788;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE HCV polyprotein (Genome polyprotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96362158; PubMed=8720135;  
 RA Seki M., Honda Y.;  
 RT "Phosphotyrosine antisense oligodeoxynucleotides capable of  
 RT inhibiting Hepatitis C virus gene expression: In vitro translation  
 RT assay";  
 RL J. Blochem. 118:1199-1204 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL; D45172; BAA08120.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00329; P00329.  
 DR HSSP; P26663; LUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_tyrp\_in.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR007451; HCV NS4A.  
 DR InterPro; IPR001490; HCV NS4B.  
 DR InterPro; IPR002868; HCV NS5A.  
 DR InterPro; IPR002166; HCV NS5B.  
 DR InterPro; IPR001650; HCV RdRp.  
 DR InterPro; IPR004109; Peptidase C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_BS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVIR.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00398; viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326880 MW; EED440E6A050E766 CRC64;

Query Match 96.0%; Score 1970; DB 12; Length 3010;  
 Best Local Similarity 94.1%; Pred. No. 1.2e-155;  
 Matches 370; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1 MASCGAVFGLALITSPYKVLARLIMVLOYLITRVEAHQWIPPLNVRGGRDAI 60  
 DB 814 MASCGAVFGLVLLITSPYKVLAKLIMVLOYLITRVEAHQWIPPLNVRGGRDAI 873  
 QY 61 ILLTAHPHLLFDITKLLALFGLPMLVQAGITVVPVFAQGLIRACMLVKAAGHY 120  
 DB 874 ILLTAHPHLLFDITKLLALFGLPMLVQAGITVVPVFAQGLIRACMLVKAAGHY 933  
 QY 121 VGMAMKLAALTCGVVDHLPLODMAHAGLDLAVANPEVFSMEVKIITWGAADTAAC 180  
 DB 934 VGMAMKLAALTCGVVDHLPLODMAHAGLDLAVANPEVFSMEVKIITWGAADTAAC 993  
 QY 181 GDIISGLPVASARGREILLGPADNFGQGWRLAEITAYSQGTRELICITSLTGRDKN 240  
 DB 994 GDIISGLPVASARGREILLGPADNFGQGWRLAEITAYSQGTRELICITSLTGRDKN 1053  
 QY 241 QVEGEVQVSTAQSFATCTVNGVCMVTHGAGAKTLAGKPGITQMTNTNVDDLVGMOS 300  
 DB 1054 QVEGEVQVSTAQSFATCTVNGVCMVTHGAGAKTLAGKPGITQMTNTNVDDLVGMOS 1113  
 QY 301 PGARSMPTCTCGSSDLYLVTRHADVIVPRRRGDSRGSLLSPVSYLKSSGGPLLCPS 360  
 DB 1114 PGARSLTPTCTCGSSDLYLVTRHADVIVPRRRGDSRGSLLSPVSYLKSSGGPLLCPS 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 9  
 Q81755  
 ID Q81755 PRELIMINARY; PRT; 1186 AA.  
 AC Q81755;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91088550; PubMed=2175903;  
 RA Kato N., Hijioka M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RT "Molecular cloning of the human hepatitis C virus genome from Japanese  
 RT patients with non-A, non-B hepatitis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528 (1990).  
 CC -1-  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00329; P00329.  
 DR HSSP; P26663; LUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_tyrp\_in.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR007451; HCV NS4A.  
 DR InterPro; IPR001490; HCV NS4B.  
 DR InterPro; IPR002868; HCV NS5A.  
 DR InterPro; IPR002166; HCV NS5B.  
 DR InterPro; IPR001650; HCV RdRp.  
 DR InterPro; IPR004109; Peptidase C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_BS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVIR.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.

RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704 (1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91140698; PubMed=1847440;  
 RA Takamizawa A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E.,  
 RA Andoh T., Yoshida I., Okayama H.,  
 RT "The structure and organization of the Hepatitis C virus genome  
 RT isolated from human carriers.";  
 RL J. Virol. 65:1105-1113 (1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.,  
 RA Bradley D.W., Kuo G., Houghton M.,  
 RT "Genetic organization and diversity of the hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455 (1991).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230206; PubMed=1331449;  
 RA Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.,  
 RT "The Taiwanese hepatitis C virus genome: Sequence determination and  
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113 (1992).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230232; PubMed=13314459;  
 RA Okamoto H., Kuzai K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishiro S.,  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: Comparative study of four distinct  
 RT genotypes.";  
 RL Virology 188:331-341 (1992).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93323208; PubMed=8392606;  
 RA Hijikata M., Mizushima H., Akagi T., Mori S., Kakinuchi N., Kato N.,  
 RA Tanaka T., Kimura K., Shimotohno K.,  
 RT "Two distinct proteinase activities required for the processing of a  
 RT putative nonstructural precursor protein of hepatitis C virus.";  
 RL J. Virol. 67:4665-4675 (1993).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Hijikata M.,  
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: D11397; BAA20975.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: PS0329; PS0329.  
 DR PDB: 1DXF; 28-MAR-02.  
 DR GO: GO:0005524; P:ATP binding; IEA.  
 DR GO: GO:0008026; P:ATP dependent helicase activity; IEA.  
 DR GO: GO:0016787; F:Hydrolase activity; IEA.  
 DR GO: GO:0003676; F:Nucleic acid binding; IEA.  
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO: GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO: GO:0019087; P:Viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_\_trypsin.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR ATP-binding; Helicase; Hydrolase.  
 FT NON TER 1  
 SQ SEQUENCE 1186 AA; 126280 MW; 34170478BA23729A CRC64;

Query Match 95.9%; Score 1968; DB 12; Length 1186;  
 Best Local Similarity 94.4%; Pred. No. 5.2e-156;  
 Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MAASGCAVFIGLALLTLSPYKTLARLIMWLQYLITRYEALQVWIPPLNVRGSDAI 60  
 DB 92 MAASGCAVFIGLALLTLSPYKTLARLIMWLQYLITRYEALQVWIPPLNVRGSDAI 151  
 QY 61 ILITCAVPELIPDITKLLALISPLAVLQAGITKVPFPAQGLIACMLVRRAGGHY 120  
 DB 152 ILITCAVPELIPDITKLLALISPLAVLQAGITKVPFPAQGLIACMLVRRAGGHY 211  
 QY 121 VQMAFMKLAALGTGVVDHLTPLQDMAHAGRLDAVAVEPIFSDMEVKIITMGADTAAC 180  
 DB 212 VQMAFMKLAALGTGVVDHLTPLQDMAHAGRLDAVAVEPIFSDMEVKIITMGADTAAC 271  
 QY 181 GDIIISGLPVSARREKILLGPADEFGQWELLAPITAVSQOTRGLICITISLTGRDXN 240  
 DB 272 GDIIISGLPVSARREKILLGPADEFGQWELLAPITAVSQOTRGLICITISLTGRDXN 331  
 QY 241 QVEGEVQVNSTATQSFLATCNVGCWTFVHGAGSKTLAGPKPTQMTYTNVDDLVGMOA 300  
 DB 332 QVDEGVQVNSTATQSFLATCNVGCWTFVHGAGSKTLAGPKPTQMTYTNVDDLVGMPA 391  
 QY 301 PPGARSMTPTCGSSSDIYLVTRHADVIPIRRRGDSRSLSLSPREVSYLKSSSGGFLCPS 360  
 DB 392 PPGARSMTPTCGSSSDIYLVTRHADVIPIRRRGDSRSLSLSPREVSYLKSSSGGFLCPS 451  
 QY 361 GHANGIFPAANTCGVAKANDFIVESEMETMR 393  
 DB 452 GHANGIFPAANTCGVAKANDFIVESEMETMR 484  
 RESULT 10  
 ID Q81817 PRELIMINARY; PRT; 2284 AA.  
 AC Q81817;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein precursor (Genome polypeptide).  
 OS Hepatitis C virus type 2.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=40271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94068484; PubMed=7504283;  
 RA Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,  
 RA Akagi T., Kimura K., Shimotohno K.,  
 RT "Proteolytic processing and membrane association of putative  
 RT nonstructural proteins of hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94333910; PubMed=8056334;  
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.,  
 RT "Identification of the domain required for trans-cleavage activity of  
 RT hepatitis C viral serine proteinase.";  
 RL Gene 145:215-219 (1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95056078; PubMed=7966638;  
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.,  
 RT "Hepatitis C virus polyprotein processing: kinetics and mutagenic  
 RT analysis of serine proteinase-dependent cleavage.";  
 RL J. Virol. 68:8418-8422 (1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95156583; PubMed=7853491;  
 RA Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.,  
 RT "Hepatitis C virus-encoded nonstructural protein NS4a has versatile

RT functions in viral protein processing.";  
 RL J. Virol. 69:1575-1581(1995).  
 DR EMBL; D16435; BAA03905.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P50329; P50329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0019012; C:violin; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008226; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR00745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01538; HCV\_NS3; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KM Nonstructural protein; Polypotein; RNA-directed RNA polymerase;  
 KM Signal; Transferase.  
 FT SIGNAL 55 83 20  
 FT CHAIN 21 54  
 FT CHAIN 84 300  
 FT CHAIN 301 931  
 FT CHAIN 932 985  
 FT CHAIN 986 1246  
 FT CHAIN 1247 1693  
 FT CHAIN 1694 2284  
 FT CHAIN 2284 AA; 247213 MW; DC272A1517046337 CRC64;  
 SQ SEQUENCE  
 Query March 95.9%; Score 1968; DB 12; Length 2284;  
 Best Local Similarity 94.4%; Pred. No. 1.2e-155;  
 Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 301 PGASMTPECTGSSDLVYTRHADVIPRRRGDSRGLLSRPVSYLKSGSGGILLCPSS 360  
 DB 388 PGASMTPECTGSSDLVYTRHADVIPRRRGDSRGLLSRPVSYLKSGSGGILLCPSS 447  
 QY 361 GHVAGIFRAAVCTRGVAKAVDFIPVESMETTR 393  
 DB 448 GHVAGIFRAAVCTRGVAKAVDFIPVESMETTR 480  
 RESULT 11  
 ID P89966 PRELIMINARY; PRT; 3010 AA.  
 AC P89966;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE RNA for Polypotein (Genome polypotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-type 1b;  
 RA Tanaka T.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-type 1b;  
 RA TANAKA T.;  
 RT "TMORF.";  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL; P89872; BAA14035.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00804; P00804.  
 DR PIR; P50329; P50329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019029; C:viral capsid; IEA.  
 DR GO; GO:0018031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002518; HCV\_capsid.  
 DR InterPro; IPR002519; HCV\_core.  
 DR InterPro; IPR002511; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5b.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.

Qy	1	MAASCGGAVFPGIALTLSPYKYLARLLMTLCYLITRVAHQQWIPPLNATGGRAI	60
Db	814	MAASCGGAVFPGIALTLSPYKYLARLLMTLCYLITRVAHQQWIPPLNATGGRAI	873
Qy	61	ILITCAVHEPILPITKLALFEGPLNVLQAGITKRVYFYRAQGLIRACMLVRKAAGHY	120
Db	874	ILITCAVHEPILPITKLALFEGPLNVLQAGITKRVYFYRAQGLIRACMLVRKAAGHY	933
Qy	121	VQMAFKKALALTGYVYVDHLTPLEDDAHNAGIRDLAVAVEPILTSDDMEKILITWQADTAA	180
Db	934	VQMALVTKALALTGYVYVDHLTPLEDDAHNAGIRDLAVAVEPILTSDDMEKILITWQADTAA	993
Qy	181	GDIIISGLPVASRRREKILIGPADNFEGQWRLLAPITAVSQQTGILGCIITSLTGDRKN	240

```

Db 994 GDIIILGIPVARSARGRELLIPADSFEGQGRILLAPITAVYQOTRGLGCIITSLTGRDXN 1053
Qy 241 QVEGEVQVNSTATQSFATCVNGVCWTFVHGAGSKTLGKGPITQMTYNVDQDLVGMQA 300
Db 1054 QVEGEVQVNSTATQSFATCVNGVCWTFVHGAGSKTLGKGPITQMTYNVDQDLVGMQA 1113
Qy 301 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGLSPRPVSLKSGSGPLCP 360
Db 1114 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGLSPRPVSLKSGSGPLCP 1173
Qy 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 393
Db 1174 GHVVGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 13
Q99AU2 PRELIMINARY; PRT; 3010 AA.
AC Q99AU2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Genome polypeptide.
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=chimera of HCV-BK;
RA Thomson M., Nasctiment M., Gonzales S., Murthy K., Rehmann B.,
RA Liang J.;
RT "Analyses of viral sequences and virus-specific immune responses
RT during serial passage of an infectious hepatitis C virus serotype 1b
RT clone in chimpanzees";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases
CC -1- SUBMITT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN C AND RNA (BY SIMILARITY).
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL: AF333324; AAK0509.1; -.
DR PIR: A61196; A61196.
DR PIR: P00246; P00246.
DR PIR: P00804; P00804.
DR PIR: P00329; P00329.
DR HSSP: P26663; INS3.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F: electron transporter activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0016740; F: transferase activity; IEA.
DR GO: GO:0006118; P: electron transport; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019079; P: viral genome replication; IEA.
DR GO: GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR009003; Cys Ser tyrosin.
DR InterPro: IPR000345; CytC_heme_B8.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR005222; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002516; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.

```

```

DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVit.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01539; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02807; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC_1.
DR PROSITE: PS00190; CYTOCHROME C; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327007 MW; 053B9A653B0AB335 CRC64;

Query Match 95.8%; Score 1967; DB 12; Length 3010;
Best Local Similarity 93.9%; Pred. No. 2.1e-155;
Matches 369; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 1 NAASCGAVFGLALITLSPYKVLARLLIWLQYLITRVEAHLQWIPPLNVRGGRDAI 60
Db 814 NAASCGAVFGLVLTLSFYKVFARLLIWLQYFIRAEHLQWVPLNVRGGRDAI 873
Qy 61 ILLTCVHPELFDITKLLAFGLMVLQAGITVFPVAGGILRACMLVRKAGGHY 120
Db 874 ILLTCVHPELFDITKLLAFGLMVLQAGITVFPVAGGILRACMLVRKAGGHY 933
Qy 121 VQMAFMKLAALGTGVYDHLFPLQMAHAGLDLAAVEPVYFSDEVKIITMGADTAAC 180
Db 934 VQMAFMKLAGLGTIYIYNHLPLRDMAHAGLDLAAVEPVYFSDEVKIITMGADTAAC 993
Qy 181 GIIIGLPSVARSARGREILLGPADNFEQGRILLAPITAVYQOTRGLGCIITSLTGRDXN 240
Db 994 GIIIGLPSVARSARGREILLGPADNFEQGRILLAPITAVYQOTRGLGCIITSLTGRDXN 1053
Qy 241 QVEGEVQVNSTATQSFATCVNGVCWTFVHGAGSKTLGKGPITQMTYNVDQDLVGMQA 300
Db 1054 QVEGEVQVNSTATQSFATCVNGVCWTFVHGAGSKTLGKGPITQMTYNVDQDLVGMQA 1113
Qy 301 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGLSPRPVSLKSGSGPLCP 360
Db 1114 PGGASMTPTCTGSSDLYLVTRHADVPVRRGDSRGLSPRPVSLKSGSGPLCP 1173
Qy 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 393
Db 1174 GHVVGIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 14
Q99TD7 PRELIMINARY; PRT; 3010 AA.
AC Q99TD7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Genome polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV217;
RA Takahashi K., Ohta Y., Kanai K., Baba K., Hijioka M.,
RA Hatahara T., Ohta Y., Kanai K., Baba K., Hijioka M.,
RA Mishiro S.;

```

"Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 'progression score' revisited." RT  
 Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases. RL  
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A CC  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: CC  
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF CC  
 PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: AB049100; BAB1813.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: PQ0804; PQ0804.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26663; IXP.  
 DR MEROPS: S29.002; -.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; F: electron transport; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; Cyt\_heme\_35.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR PRODOM: PD16062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SO SEQUENCE 3010 AA; 326909 MW; 5505C62EH0DA0519 CRC64;

Query Match 95.8%; Score 1967; DB 12; Length 3010;  
 Best Local Similarity 94.4%; Pred. No. 2.1e-155;  
 Matches 371; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 1 MAASCGAVGIGLALTLSPYKYLARLIMWLYLTRVEAHLYQWIPELNVRGGDAI 60  
 DQ 814 VAASCGAVGIGLALTLSPYKYLARLIMWLYLTRVEAHLYQWIPELNVRGGDAI 873  
 QY 61 ILTCAVHPELIPITLLAIGPLVLAQGITKPYFRACGLRACMLVRKAGSHY 120

DB 874 ILTCAVHPELIPITLLAIGPLVLAQGITRVEYFVPAQGLIRACMLVRKAGSHY 933  
 QY 121 VQAMFMKLAALGTYYVDHLPLQDMNAGRLDAVVEPIFSDMEXKITMGADTAAC 180  
 DB 934 IQMLVLAALALGTYVDHLAPLQDMNAGRLDAVVEPIFSDMEXKITMGADTAAC 993  
 QY 181 GDIISGLPVSARRREILGPADNFEQGWMLLAPITAYSQQTRGLIGCIITSLTGDKN 240  
 DB 994 GDIIILGFLVSARRREILGPADNFEQGWMLLAPITAYSQQTRGLIGCIITSLTGDKN 1053  
 QY 241 QVEGEVQVSTATOSPLATCNGVCWTFVHGASGKTLTAGPGPTQMTYTNVDDLVGMOA 300  
 DB 1054 QVEGEVQVSTATOSPLATCNGVCWTFVHGASGKTLTAGPGPTQMTYTNVDDLVGMOA 1113  
 QY 301 PFGARSMTPCTCGSSDLYLTRHADVIPIRRRGDSRSLSPREVSYLKSGSGPLLCPS 360  
 DB 1114 PFGARSMTPCTCGSSDLYLTRHADVIPIRRRGDSRSLSPREVSYLKSGSGPLLCPS 1173  
 QY 361 GHANGIFPAANCTGVAKAVDFIVESEMETMR 393  
 DB 1174 GHANGIFPAANCTGVAKAVDFIVESEMETMR 1206

RESULT 15  
 Q9QIX6 PRELIMINARY; PRT; 3010 AA.  
 ID Q9QIX6;  
 AC Q9QIX6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polypotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD8-1;  
 RX MEDLINE=2001325; PubMed=10544098;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,  
 RA Tazawa J.I., Izumi N., Marumo F., Sato C.;  
 RT "Time-related changes in full-length hepatitis C virus and hepatitis  
 RT activity.";  
 RL Virology 263:244-253 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD8-1;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,  
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AF165059; AAD56194.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: PQ0246; PQ0246.  
 DR PIR: PQ0254; PQ0254.  
 DR PIR: PQ0804; PQ0804.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26663; IXP.  
 DR MEROPS: S29.002; -.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.



DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_cryptin.  
 DR InterPro; IPR00410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF002271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR Pfam; PF0186062; HCV\_NS1; 1.  
 DR PRODOM; PD186062; DEXdc; 1.  
 DR SMART; SM00487; DEXdc; 1.  
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327466 MW; 4613744EC4DA013 CRC64;

Query Match 95.8%; Score 1967; DB 12; Length 3010;  
 Best Local Similarity 94.4%; Pred. No. 2.1e-155;  
 Matches 371; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 1 MAASCGAVFIQALTLSPYKVLARLIMLOYLITRVEAHLOVWIPPLNVRGGRDAI 60  
 DB 814 MAASCGAVFIQALTLSPYKVLARLIMLOYLITRVEAHLOVWIPPLNVRGGRDAI 873  
 QY 61 ILITCAVHPELIPDITKLLAIFGPLMVLQAGITKVPYFVRAOGLIRACMLVRRAAGHY 120  
 DB 874 ILITCAVHPELIPDITKLLAIFGPLMVLQAGITKVPYFVRAOGLIRACMLVRRAAGHY 933  
 QY 121 VQMAFMFLAALTGYVVDHLPLQDMAHAGLRDLAAVEVPIPSDMEVKIITWGADTAAC 180  
 DB 934 VQMAFMFLAALTGYVVDHLPLQDMAHAGLRDLAAVEVPIPSDMEVKIITWGADTAAC 993  
 QY 181 GDIIISGLPVSAARRGREILLGPADNFEQGWRLAPITAYSQOTRGLIGCIITSLTGRDKN 240  
 DB 994 GDIIISGLPVSAARRGREILLGPADNFEQGWRLAPITAYSQOTRGLIGCIITSLTGRDKN 1053  
 QY 241 QVEGEVQVSTATQSFATCTGAGVCTVFFGAGSKITLAGKGPITQYTYTNVDQDLVGMQA 300  
 DB 1054 QVEGEVQVSTATQSFATCTGAGVCTVFFGAGSKITLAGKGPITQYTYTNVDQDLVGMQA 1113  
 QY 301 PGARSMTPCTCGSSDLYLTVRAHDAVIVRRRGDSRSLSPRPVSYLKSSSGGPLLCPG 360  
 DB 1114 PGARSMTPCTCGSSDLYLTVRAHDAVIVRRRGDSRSLSPRPVSYLKSSSGGPLLCPG 1173  
 QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
 DB 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

Search completed: May 6, 2004, 09:35:46  
 Job time : 37.7566 secs



Fri May 7 13:37:00 2004

us-10-650-585-11.ra1

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004. CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 15.4246 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALTLSP.....RGVAKAVDFIPVESMETTR 393

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/6C\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/6D\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1951	95.0	2201	4	US-09-539-601-6
2	1951	95.0	2201	4	US-09-539-601-15
3	1951	95.0	3010	4	US-09-539-601-3
4	1951	95.0	3010	4	US-09-539-601-21
5	1951	95.0	3010	4	US-09-539-601-27
6	1946	94.8	1692	3	US-09-263-933-4
7	1946	94.8	1692	3	US-09-919-901-4
8	1946	94.8	2307	3	US-09-263-933-2
9	1946	94.8	2307	3	US-09-919-901-2
10	1944	94.7	3010	4	US-09-539-601-13
11	1943	94.6	1692	3	US-09-263-933-11
12	1943	94.6	1692	4	US-09-919-901-11
13	1943	94.6	2307	3	US-09-263-933-9
14	1943	94.6	2307	3	US-09-919-901-9
15	1934	94.2	1692	3	US-09-263-933-18
16	1934	94.2	1692	4	US-09-919-901-18
17	1934	94.2	2307	3	US-09-263-933-16
18	1934	94.2	2307	3	US-09-919-901-16
19	1928	93.9	3010	3	US-09-014-416-3
20	1888	92.0	2013	1	US-08-324-977-12
21	1888	92.0	2013	1	US-08-384-616-12
22	1888	92.0	2013	2	US-08-904-686A-12
23	1888	92.0	2013	3	US-09-315-850-12
24	1888	92.0	2201	4	US-08-952-981A-2
25	1888	92.0	2620	1	US-08-324-977-12
26	1888	92.0	2620	2	US-08-384-616-32
27	1888	92.0	2620	2	US-08-904-686A-32

28	1888	92.0	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1888	92.0	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1888	92.0	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1888	92.0	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1888	92.0	2621	2	US-09-315-850-36	Sequence 36, Appl
33	1888	92.0	3010	1	US-08-324-977-12	Sequence 2, Appl
34	1888	92.0	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1888	92.0	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1888	92.0	3010	2	US-08-904-686A-2	Sequence 14, Appl
37	1888	92.0	3010	2	US-08-904-686A-14	Sequence 14, Appl
38	1888	92.0	3010	2	US-09-315-850-2	Sequence 2, Appl
39	1888	92.0	3010	3	US-09-315-850-14	Sequence 14, Appl
40	1888	92.0	3010	3	US-08-188-281B-12	Sequence 12, Appl
41	1767	86.1	1648	5	PCT-US94-07280-12	Sequence 12, Appl
42	1767	86.1	1648	5	PCT-US95-01087-12	Sequence 12, Appl
43	1767	86.1	3011	1	US-08-188-281B-1	Sequence 1, Appl
44	1767	86.1	3011	1	US-08-453-552-1	Sequence 1, Appl
45	1767	86.1	3011	1	US-08-453-552-1	Sequence 1, Appl

# ALIGNMENTS

RESULT 1  
US-09-539-601-6  
; Sequence 6, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartschlag, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match 95.0%; Score 1951; DB 4; Length 2201;  
Best Local Similarity 93.1%; Pred. No. 8.8e+187;  
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY	1	MAASCGAVFGLALTLSPYKVLARLIMWLQYLITRVEAHLQWIPPLNVRGGRDAI	60
DB	5	MAASCGAVFGLALTLSPHYKFLARLIMWLQYFIRAEHLQWIPPLNVRGGRDAV	64
QY	61	ILITCAVPEELFDITKLLALFGLPLMTQAGITVPPVRAQGLIRACMLYRKAAGHY	120
DB	65	ILITCAVPEELFDITKLLALFGLPLMTQAGITVPPVRAHGLIRACMLYRKAAGHY	124
QY	121	VQMAEFKLLALGTYYVDHLTFLQDMAGLRDLAAVEPVIFSDMEVKIITWGADTAAC	180
DB	125	VQMALMKLALAGTYVDHLTFLRDMAGLRDLAAVEPVIFSDMETVIMWGADTAAC	184
QY	181	GDIISGLPVASARGREILIGPNDFFGQGRLLAPITAYSOQTRGLGCIITSLGRDN	240
DB	185	GDIISGLPVASARGREILIGPNDFFGQGRLLAPITAYSOQTRGLGCIITSLGRDN	244
QY	241	QVEGEVQVSTAGSFATCVNGVCMTHFGAGSKTLGPKPITOMYTNVDQDVLGQA	300
DB	245	QVEGEVQVSTAGSFATCVNGVCMTHFGAGSKTLGPKPITOMYTNVDQDVLGQA	304
QY	301	PPGARSMTCTCGSSDLYLVTRHADVIPYRRGDSRGSILSPVSYLKGSSGGPLCP	360
DB	305	PPGARSMTCTCGSSDLYLVTRHADVIPYRRGDSRGSILSPVSYLKGSSGGPLCP	364
QY	361	GHAAGIFRAAVCTRGVAKAVDFIPVESMETTR	393

Db 365 GHAVGIFRAVCTRGVAKAVDFVPVESMETTMR 397

```

RESULT 2
US-09-539-601-15
; Sequence 15, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Batteneschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-10
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-539-601-15

```

Query Match	95.0%	Score 1951,	DB 4;	Length 2201,
Best Local Similarity	93.1%;	Pred. No. 8.8e-187;		
Matches 366;	Conservative 14;	Mismatches 13;	Indels 0	

QY	MAASCGAAVEIGALATLTSPPYKVLARILIMLOXYLIRVEAHLOVMPILNAVNGSDAI	60
Db	5 MAASCGAAVFGILITLTSPPHYKFLARILIMLOXYLIRABAHLOVMPILNAVNGSDAV	64
QY	1LTTCANHEBELLFDITKLLALFGLPMVLQAGITRVYFVBAQGLIRACMLVRKAAGHY	120
Db	65 ILTTCANHEBELLFTITKLLALLGLPMVLQAGITKVPFVBAHGLIRACMLVRKAAGHY	124
QY	121 VQNAFKMLALIGTYIYDHLTPLOMANAHLRLDANVEPIFSDMEKIIITWADNPAAC	180
Db	125 VQNAFKMLALIGTYIYDHLTPLRMANAHLRLDANVAEPVFSDMETKVIITWADNPAAC	184
QY	181 GDITISGLPVSARSGREIILGPADNFEQGOCMRLAPITAYSOOTSGILGCIITSLTGSDKN	240
Db	185 GDITISGLPVSARSGREIILGPADNFBGGOCMRLAPITAYSOOTSGILGCIITSLTGSDKN	244
QY	241 QVEGEVQVNSTATOSFLATCVNGVCMVTFHAGSKTLTAGPKPIITOMYTNVDQDLVMOA	300
Db	245 QVEGEVQVNSTATOSFLATCVNGVCMVTFYHAGSKTLTAGPKPIITOMYTNVDQDLVMOA	304
QY	301 PPGARSMTPTCGSSSDLYLYTHADVITPYRRRGDSRGLLSPPRVSYLKGSGGPPLLCPB	360
Db	305 PPGARSMTPTCGSSSDLYLYTHADVITPYRRRGDSRGLLSPPRVSYLKGSGGPPLLCPB	364
QY	361 GHAUVGFRAAVCTRGYAKAVDFIPVESMETMNR	393
Db	365 GHAUVGFRAAVCTRGYAKAVDFIPVESMETMNR	397

RESULT 3  
US-09-539-601-3  
; Sequence 3, Application US/09539601C

ORGANISM: Hepatitis C virus  
US-09-539-601-3

Query Match	95.0%	Score 1951;	DB 4;	Length 3010;
Best Local Similarity	93.1%	Pred. No. 1.4e-186;		
Matches 366;	Conservative 14;	Mismatches 13;	Indels 0;	Gaps 0;

QY	1	MAASCGAVFGLIALTLSPYKTLARLIMVQYLITREALQVILPILNVRGDAI	60
Db	814	MAASCGAVFGLIILTLSPHYKFLRLIMVQYITREAHQVILPILNVRGDAV	873
QY	61	ILLTCAVHPELIFDITKLLAIFGLNVLQAGITKVPYFVAGLLRACMLVRKAGHY	120
Db	874	ILLTCAIHPELIFTITKLLALGLPMLVLAQGITKVPYFVAGLLRACMLVRKAGHY	933
QY	121	VQMAFKLAALTGYVDHLTLPDQMAAGRLDAVAEBVIFSDMEVKITWGADTAAC	180
Db	934	VQMAFMKLAALTGYVDHLTPLRDMMAAGRLDAVAEBVIFSDMEVKITWGADTAAC	993
QY	181	GLIIGLPSARRGELILSPADNEQGRLLAPITAYSQQRGLGCIITSLGRDKN	240
Db	994	GIIILGLPSARRGELIHLGPADSLEQGRLLAPITAYSQQRGLGCIITSLGRDRN	1053
QY	241	QVEGEVQVSTATQSLFATLCVNGVCMVFPAGASKTLAGKGIITQMTYNNVDLWQQA	300
Db	1054	QVEGEVQVSTATQSLFATLCVNGVCMVYHAGASKTLAGKGIITQMTYNNVDLWQQA	1113
QY	301	PGAGASMPPTCGSSDLYLTVRAADVIVERRGDSRGLSPPVSYLKGSSGGPLLCP	360
Db	1114	PGASSTLPCTCGSSDLYLTVRAADVIVRRRGDSRGLSPPVSYLKGSSGGPLLCP	1173
QY	361	GHAVGIFRAAVCTRGVAKAVDFIVESMEETMR	393
Db	1174	GHAVGIFRAAVCTRGVAKAVDFIVESMEETMR	1206

```

RESULT 4 -
US-09-539-601-21
Sequence 21, Application US/09539601C
Patent No. 6630343
GENERAL INFORMATION:
APPLICANT: Bartschschlager, Ralf FM
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ. ID NOS: 51
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 3010
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-21

```

Query Match	95.0%;	Score 1951;	DB 4;	Length 3010;
Best Local Similarity	93.1%;	Pred. No. 1.4e-18;		
Matches 366;	Conservative 14;	Mismatches 13;	Indels 0;	Gaps 0;

[illegible]

```

Db      994 GDIILGLPVSARGRGRIHIGPADSLEGQGWRLAPTAASQOTRGLGCIITSLTGRDRN 1053
Qy      241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSGKTLGPKPITQMTYTNVDQDLVGMQA 300
Db      1054 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSGKTLGPKPITQMTYTNVDQDLVGMQA 1113
Qy      301 PGARSMPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGLSPRPVSYLKSGSGGPLLCP 360
Db      1114 PGARSLTPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGLSPRPVSYLKSGSGGPLLCP 1173
Qy      361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393
Db      1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 1206

```

## RESULT 5

```

US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FM
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-539-601-27

```

Query Match 95.0%; Score 1951; DB 4; Length 3010;

Best Local Similarity 93.1%; Pred. No. 1,4e-186; Matches 365; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```

Qy      1 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQWIPPLNARGSDAI 60
Db      814 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQWIPPLNARGSDAY 873
Qy      61 ILLTCAVHPELFDITKLLAIFGPIVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120
Db      874 ILLTCAVHPELFDITKLLAIFGPIVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 933
Qy      121 VQMAFMKLAALTYGVYDHLTPLODMAHGRLDAVAEPIVPSDMEVKIITWGADTAAC 180
Db      934 VQMAFMKLAALTYGVYDHLTPLODMAHGRLDAVAEPIVPSDMEVKIITWGADTAAC 993
Qy      181 GDITISGLPVARSRRREILLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDN 240
Db      994 GDITISGLPVARSRRREILLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDN 1053
Qy      241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSGKTLGPKPITQMTYTNVDQDLVGMQA 300
Db      1054 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSGKTLGPKPITQMTYTNVDQDLVGMQA 1113
Qy      301 PGARSMPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGLSPRPVSYLKSGSGGPLLCP 360
Db      1114 PGARSLTPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGLSPRPVSYLKSGSGGPLLCP 1173
Qy      361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393
Db      1174 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 1206

```

## RESULT 6

```

US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940

```

```

; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1998-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; US-09-263-933-4

```

Query Match 94.8%; Score 1946; DB 3; Length 1692;

Best Local Similarity 92.9%; Pred. No. 1.9e-186; Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

```

Qy      1 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQWIPPLNARGSDAI 60
Db      93 MAASCGAVFTGLALTLSPYKVLARLIWLOYLITRVAHLQWIPPLNARGSDAI 152
Qy      61 ILLTCAVHPELFDITKLLAIFGPIVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120
Db      153 ILLTCAVHPELFDITKLLAIFGPIVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 212
Qy      121 VQMAFMKLAALTYGVYDHLTPLODMAHGRLDAVAEPIVPSDMEVKIITWGADTAAC 180
Db      213 VQMAFMKLAALTYGVYDHLTPLODMAHGRLDAVAEPIVPSDMEVKIITWGADTAAC 272
Qy      181 GDITISGLPVARSRRREILLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDN 240
Db      273 GDITISGLPVARSRRREILLGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDN 332
Qy      241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSGKTLGPKPITQMTYTNVDQDLVGMQA 300
Db      333 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSGKTLGPKPITQMTYTNVDQDLVGMQA 392
Qy      301 PGARSMPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGLSPRPVSYLKSGSGGPLLCP 360
Db      393 PGARSLTPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGLSPRPVSYLKSGSGGPLLCP 452
Qy      361 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 393
Db      453 GHAVGIFRAAVCTRGVAKAVDPFVPSMETTMR 485

```

## RESULT 7

```

US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6539738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; EARLIER FILING DATE: 2001-08-02
; EARLIER APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692

```

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-4

Query Match 94.8%; Score 1946; DB 4; Length 1692;  
Best Local Similarity 92.9%; Pred. No. 1.9e-186;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLOYLITRVEAHLQVWIPPLNVRGGRDAI 60  
DB 93 MAASCGAVFGLVLLTSPYKVLARLIWLOYLITRVEAHLQVWIPPLNVRGGRDAI 152  
QY 61 ILTCAVHPELIPDITKLLAIFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 120  
DB 153 ILMCAVHPELIPDITKLLAIFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 212  
QY 121 VQMAFMKALTLGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 180  
DB 213 VQMAFMKALTLGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 272  
QY 181 GDIISGLPVARSRGREILGPADNFEQGWRLLAPITAVSQOTRGLGCIITSLTGRDXN 240  
DB 273 GDIISGLPVARSRGREILGPADNFEQGWRLLAPITAVSQOTRGLGCIITSLTGRDXN 332  
QY 241 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGPKPITQWYTNVDDLVGMOA 300  
DB 333 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGPKPITQWYTNVDDLVGMOA 392  
QY 301 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRRGRDSDRLSPRVSYLKSSGGPILCP 360  
DB 393 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRRGRDSDRLSPRVSYLKSSGGPILCP 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 485

RESULT 8  
US-09-263-933-2  
Sequence 2, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
CURRENT FILING DATE: 1999-03-08  
EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-2

Query Match 94.8%; Score 1946; DB 3; Length 2307;  
Best Local Similarity 92.9%; Pred. No. 3e-186;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLOYLITRVEAHLQVWIPPLNVRGGRDAI 60  
DB 185 MAASCGAVFGLVLLTSPYKVLARLIWLOYLITRVEAHLQVWIPPLNVRGGRDAI 244  
QY 61 ILTCAVHPELIPDITKLLAIFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 120  
DB 245 ILMCAVHPELIPDITKLLAIFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 304

QY 121 VQMAFMKALTLGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 180  
DB 305 VQMAFMKALTLGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 364  
QY 181 GDIISGLPVARSRGREILGPADNFEQGWRLLAPITAVSQOTRGLGCIITSLTGRDXN 240  
DB 365 GDIISGLPVARSRGREILGPADNFEQGWRLLAPITAVSQOTRGLGCIITSLTGRDXN 424  
QY 241 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGPKPITQWYTNVDDLVGMOA 300  
DB 425 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAAGPKPITQWYTNVDDLVGMOA 484  
QY 301 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRRGRDSDRLSPRVSYLKSSGGPILCP 360  
DB 485 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRRGRDSDRLSPRVSYLKSSGGPILCP 544  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 577

RESULT 9  
US-09-919-901-2  
Sequence 2, Application US/09919901  
Patent No. 6599738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-2

Query Match 94.8%; Score 1946; DB 4; Length 2307;  
Best Local Similarity 92.9%; Pred. No. 3e-186;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLOYLITRVEAHLQVWIPPLNVRGGRDAI 60  
DB 185 MAASCGAVFGLVLLTSPYKVLARLIWLOYLITRVEAHLQVWIPPLNVRGGRDAI 244  
QY 61 ILTCAVHPELIPDITKLLAIFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 120  
DB 245 ILMCAVHPELIPDITKLLAIFGLPLVLOAGITKVPYFVRAQGLIRACMLYKKAAGHY 304  
QY 121 VQMAFMKALTLGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 180  
DB 305 VQMAFMKALTLGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDEVKIITWGADTAAC 364  
QY 181 GDIISGLPVARSRGREILGPADNFEQGWRLLAPITAVSQOTRGLGCIITSLTGRDXN 240  
DB 425 GDIISGLPVARSRGREILGPADNFEQGWRLLAPITAVSQOTRGLGCIITSLTGRDXN 484  
QY 301 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRRGRDSDRLSPRVSYLKSSGGPILCP 360  
DB 485 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRRGRDSDRLSPRVSYLKSSGGPILCP 544  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 577

```

QY 301 PGARSMCTCGSSDLYLVTBHADVIVRRRGDSGSLSPRPVYLKSGSGPILCP 360
DB 485 PGARSLPTCTCGSSDLYLVTBHADVIVRRRGDSGSLSPRPVYLKSGSGPILCP 544
QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 393
DB 545 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 577

```

```

RESULT 10
US-09-539-601-33
; Sequence 33; Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Batten Schlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-33

```

```

Query Match 94.7%; Score 1944; DB 4; Length 3010;
Best Local Similarity 92.8%; Pred. No. 7.1e-186;
Matches 365; Conservative 14; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 1 MAASCGAVFGLALITLSPYKVLARLWLOYLITRVEAHQWIPPLNVRGGRDAI 60
DB 814 MAASCGAVFGLALITLSPYKVLARLWLOYLITRVEAHQWIPPLNVRGGRDAV 873
QY 61 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 120
DB 874 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 933
QY 121 VQAMFKLAALTGYVYVHNLPLQDMANAGLDLAVANVEPIPSDMEVKIITWGADTAAC 180
DB 934 VQAMFKLAALTGYVYVHNLPLQDMANAGLDLAVANVEPIPSDMEVKIITWGADTAAC 993
QY 181 GDIISGLPVSARRGREILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 240
DB 994 GDIISGLPVSARRGREILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 1053
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTQMTNTNDDLVGMQA 300
DB 1054 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTQMTNTNDDLVGMQA 1113
QY 301 PGARSMCTCGSSDLYLVTBHADVIVRRRGDSGSLSPRPVYLKSGSGPILCP 360
DB 1114 PGARSLPTCTCGSSDLYLVTBHADVIVRRRGDSGSLSPRPVYLKSGSGPILCP 1173
QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 393
DB 1174 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 1206

```

```

RESULT 11
US-09-263-933-11
; Sequence 11; Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A

```

```

; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-11

```

```

Query Match 94.6%; Score 1943; DB 3; Length 1692;
Best Local Similarity 92.6%; Pred. No. 3.8e-166;
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 1 MAASCGAVFGLALITLSPYKVLARLWLOYLITRVEAHQWIPPLNVRGGRDAI 60
DB 93 MAASCGAVFGLALITLSPYKVLARLWLOYLITRVEAHQWIPPLNVRGGRDAI 152
QY 61 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 120
DB 153 ILTCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVAQGLIRACMLVRKAAGHY 212
QY 121 VQAMFKLAALTGYVYVHNLPLQDMANAGLDLAVANVEPIPSDMEVKIITWGADTAAC 180
DB 213 VQAMFKLAALTGYVYVHNLPLQDMANAGLDLAVANVEPIPSDMEVKIITWGADTAAC 272
QY 181 GDIISGLPVSARRGREILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 240
DB 273 GDIISGLPVSARRGREILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDN 332
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTQMTNTNDDLVGMQA 300
DB 333 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKITLAGPKPTQMTNTNDDLVGMQA 392
QY 301 PGARSMCTCGSSDLYLVTBHADVIVRRRGDSGSLSPRPVYLKSGSGPILCP 360
DB 393 PGARSLPTCTCGSSDLYLVTBHADVIVRRRGDSGSLSPRPVYLKSGSGPILCP 452
QY 361 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 393
DB 453 GHAVGIFRAAVCTRGVAKAVDFPVESMETTR 485

```

```

RESULT 12
US-09-919-901-11
; Sequence 11; Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-11

```

```

Query Match 94.6%; Score 1943; DB 4; Length 1692;

```

Best Local Similarity 92.6%; Pred. No. 3,86-186;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

```

QY 1 MAASCGAVFGLALTLSPYKYLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
DB 93 MAASCGAVFGLVLTLSPPYKFLRLIMLQYFTTRBAHLHWIPPLNVRGGRDAI 152
QY 61 ILTCAVHPELIFDITKLLAIFGLPLMWLAGITKVFYVRAQGLIRACMLVRKAAGHY 120
DB 153 ILMCAVHPELIFDITKLLAIFGLPLMWLAGITRVPYVRAQGLIHACMLVRKAAGHY 212
QY 121 VQMAFMKLAALGTGYVVDHLTPLODMAHAGLRDLA VEPVPSDMEVKIITMGADTAAC 180
DB 213 VQMAFMKLAALGTGYVVDHLTPLODMAHAGLRDLA VEPVPSDMEVKIITMGADTAAC 272
QY 181 GDIIISGLPVSAARGREILLGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTRGDKN 240
DB 273 GDIIISGLPVSAARGREILLGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTRGDKN 332
QY 241 QVEGEVQVSTATQSFATCNGVCWTVFHGAGSKTLAGKGPITQMTYTNVDODLVGMOA 300
DB 333 QVEGEVQVSTATQSFATCNGVCWTVFHGAGSKTLAGKGPITQMTYTNVDODLVGMOA 392
QY 301 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRGDSRGLSPRPVSYLKSGSGPILCP 360
DB 393 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRGDSRGLSPRPVSYLKSGSGPILCP 452
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
DB 453 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 485

```

## RESULT 13

US-09-263-933-9  
Sequence 9, Application US/09263933  
Patent No. 6280940

```

GENERAL INFORMATION:
APPLICANT: Pottes, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/263,933
CURRENT FILING DATE: 1999-03-08
EARLIER APPLICATION NUMBER: 09/129,611
EARLIER FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 2307
TYPE: PRT
ORGANISM: Artificial Sequence
US-09-263-933-9

```

Query Match 94.6%; Score 1943; DB 3; Length 2307;  
Best Local Similarity 92.6%; Pred. No. 66-186;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

```

QY 1 MAASCGAVFGLALTLSPYKYLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
DB 185 MAASCGAVFGLVLTLSPPYKFLRLIMLQYFTTRBAHLHWIPPLNVRGGRDAI 244
QY 61 ILTCAVHPELIFDITKLLAIFGLPLMWLAGITKVFYVRAQGLIRACMLVRKAAGHY 120
DB 245 ILMCAVHPELIFDITKLLAIFGLPLMWLAGITRVPYVRAQGLIHACMLVRKAAGHY 304
QY 121 VQMAFMKLAALGTGYVVDHLTPLODMAHAGLRDLA VEPVPSDMEVKIITMGADTAAC 180
DB 305 VQMAFMKLAALGTGYVVDHLTPLODMAHAGLRDLA VEPVPSDMEVKIITMGADTAAC 364
QY 181 GDIIISGLPVSAARGREILLGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTRGDKN 240

```

```

DB 365 GDIIISGLPVSAARGREILLGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTRGDKN 424
QY 241 QVEGEVQVSTATQSFATCNGVCWTVFHGAGSKTLAGKGPITQMTYTNVDODLVGMOA 300
DB 425 QVEGEVQVSTATQSFATCNGVCWTVFHGAGSKTLAGKGPITQMTYTNVDODLVGMOA 484
QY 301 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRGDSRGLSPRPVSYLKSGSGPILCP 360
DB 485 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRGDSRGLSPRPVSYLKSGSGPILCP 544
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 577

```

## RESULT 14

US-09-919-901-9  
Sequence 9, Application US/09919901  
Patent No. 659738

```

GENERAL INFORMATION:
APPLICANT: Pottes, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 2307
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION:
US-09-919-901-9

```

Query Match 94.6%; Score 1943; DB 4; Length 2307;  
Best Local Similarity 92.6%; Pred. No. 66-186;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

```

QY 1 MAASCGAVFGLALTLSPYKYLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAI 60
DB 185 MAASCGAVFGLVLTLSPPYKFLRLIMLQYFTTRBAHLHWIPPLNVRGGRDAI 244
QY 61 ILTCAVHPELIFDITKLLAIFGLPLMWLAGITKVFYVRAQGLIRACMLVRKAAGHY 120
DB 245 ILMCAVHPELIFDITKLLAIFGLPLMWLAGITRVPYVRAQGLIHACMLVRKAAGHY 304
QY 121 VQMAFMKLAALGTGYVVDHLTPLODMAHAGLRDLA VEPVPSDMEVKIITMGADTAAC 180
DB 305 VQMAFMKLAALGTGYVVDHLTPLODMAHAGLRDLA VEPVPSDMEVKIITMGADTAAC 364
QY 181 GDIIISGLPVSAARGREILLGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTRGDKN 240
DB 365 GDIIISGLPVSAARGREILLGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTRGDKN 424
QY 241 QVEGEVQVSTATQSFATCNGVCWTVFHGAGSKTLAGKGPITQMTYTNVDODLVGMOA 300
DB 425 QVEGEVQVSTATQSFATCNGVCWTVFHGAGSKTLAGKGPITQMTYTNVDODLVGMOA 484
QY 301 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRGDSRGLSPRPVSYLKSGSGPILCP 360
DB 485 PPGARSMTPCTCGSSDLYLTVRHADVI PVRRGDSRGLSPRPVSYLKSGSGPILCP 544
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
DB 545 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 577

```



```
RESULT 15
US-09-263-933-18
; Sequence 18, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1998-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1692
; TYPE: FRT
; ORGANISM: Artificial Sequence
US-09-263-933-18
```

```
Query Match          94.2%; Score 1934; DB 3; Length 1692;
Best Local Similarity 92.4%; Pred. No. 3,1e-185;
Matches 363; Conservative 14; Mismatches 16; Indels 0; Gaps 0;
```

```
QY      1 MAASCGAVFIGALTLSPYKVLARLIWLOYLITREAHLOVWIPPLAVRGGRDAI 60
      93 MAASCGAVFVGLVLTSPYKFLARLIWLOYLITREAHLOVWIPPLAVRGGRDAI 152
      61 ILTCAVHPELIPDITKLALIFGPIWVLOAGITKVEFVFAOGLIRACMLYKKAAGHY 120
      153 ILNCAVHPELIPDITKLALIFGPIWVLOAGITKVEFVFAOGLIRACMLYKKAAGHY 212
      121 VQNAFMKLAALGTYYVYDHLTPLQDMAHAGLRDLAVAEVPIFSDEVEKITTGADTAAC 180
      213 VQNAFMKLAALGTYYVYDHLTPLQDMAHAGLRDLAVAEVPIFSDEVEKITTGADTAAC 272
      181 GDITSGIPVARGREILIGPADNFEQGWELLAPITAYSOOTRGILGCIITSLGRDKY 240
      273 GDITSGIPVARGREILIGPADNFEQGWELLAPITAYSOOTRGILGCIITSLGRDKY 332
      241 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAGPKGPIITQMTYNVQDLDVGWQA 300
      333 QVEGEVQVSTATQSFATCNGVCMVTFHAGSKTLAGPKGPIITQMTYNVQDLDVGWQA 392
      301 PPGARSMPTCTCGSSDLYLTRHADVIPIRRGDSRGLSPRVSYLTKSSGGGPIICPS 360
      393 PPGARSMPTCTCGSSDLYLTRHADVIPIRRGDSRGLSPRVSYLTKSSGGGPIICPS 452
      361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393
      453 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 485
      DB
```

Search completed: May 6, 2004, 09:39:02  
Job time : 16.4246 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 40.2025 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-11  
Perfect score: 2053  
Sequence: 1 MAASCGAVFGLALTLTSP.....RGVAKVDFPESMETTR 393

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: Published Applications AA.\*
  - 2: /cgn2\_6/prodata/1/pubppaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/prodata/1/pubppaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/prodata/1/pubppaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/prodata/1/pubppaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/prodata/1/pubppaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/prodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/prodata/1/pubppaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/prodata/1/pubppaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/prodata/1/pubppaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/prodata/1/pubppaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/prodata/1/pubppaa/US10\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/prodata/1/pubppaa/US10\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/prodata/1/pubppaa/US10C\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/prodata/1/pubppaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/prodata/1/pubppaa/US60\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/prodata/1/pubppaa/US60\_PUBCOMB.pep.\*
  - 18: /cgn2\_6/prodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2053	100.0	393	13	US-10-017-736-11
2	2053	100.0	393	13	US-10-650-585-11
3	2053	100.0	409	16	US-10-017-736-2
4	2053	100.0	409	16	US-10-650-585-2
5	1987	96.8	380	13	US-10-017-736-12
6	1987	96.8	380	13	US-10-650-585-12
7	1951	95.0	2201	13	US-10-029-807-3
8	1951	95.0	2201	13	US-10-309-561-3
9	1951	95.0	3010	12	US-10-467-000-1
10	1946	94.8	1692	14	US-09-919-901-4
11	1946	94.8	1692	14	US-10-191-966-4
12	1946	94.8	2307	14	US-09-919-901-2
13	1946	94.8	2307	14	US-10-191-966-2
14	1943	94.6	1692	10	US-09-919-801-11
15	1943	94.6	1692	14	US-10-191-966-11

16	1943	94.6	2307	10	US-09-919-901-9	Sequence 9, Appli
17	1943	94.6	2307	14	US-10-191-966-9	Sequence 9, Appli
18	1934	94.2	1692	10	US-09-919-901-18	Sequence 18, Appli
19	1934	94.2	1692	14	US-10-191-966-18	Sequence 18, Appli
20	1934	94.2	2307	10	US-09-919-901-16	Sequence 16, Appli
21	1934	94.2	2307	14	US-10-191-966-16	Sequence 16, Appli
22	1888	92.0	2201	13	US-10-085-476-2	Sequence 2, Appli
23	1842	89.7	352	13	US-10-017-736-13	Sequence 13, Appli
24	1842	89.7	352	16	US-10-650-585-13	Sequence 13, Appli
25	1778	86.6	341	13	US-10-017-736-14	Sequence 14, Appli
26	1778	86.6	2985	14	US-10-650-585-14	Sequence 14, Appli
27	1772	86.3	3011	9	US-10-259-275-40	Sequence 40, Appli
28	1766	86.0	3011	9	US-09-742-659-4	Sequence 4, Appli
29	1766	86.0	3011	10	US-09-851-894-3	Sequence 3, Appli
30	1766	86.0	3011	10	US-10-184-150-3	Sequence 3, Appli
31	1766	86.0	3011	15	US-10-128-997-3	Sequence 3, Appli
32	1766	86.0	3012	9	US-09-238-076-2	Sequence 2, Appli
33	1766	86.0	3012	10	US-09-995-937-2	Sequence 2, Appli
34	1766	86.0	3012	10	US-09-917-563-2	Sequence 2, Appli
35	1764	85.9	3011	9	US-09-916-359-2	Sequence 2, Appli
36	1764	85.9	3011	12	US-10-296-734-406	Sequence 406, App
37	1762	85.8	3011	9	US-09-238-076-20	Sequence 20, Appli
38	1762	85.8	3011	10	US-09-995-937-20	Sequence 20, Appli
39	1762	85.8	3011	10	US-09-917-563-20	Sequence 20, Appli
40	1759	85.7	2894	9	US-09-941-611-23	Sequence 23, Appli
41	1759	85.7	3011	9	US-10-044-995-23	Sequence 23, Appli
42	1759	85.7	3011	9	US-09-952-572-8	Sequence 9, Appli
43	1759	85.7	3011	9	US-09-747-419-20	Sequence 9, Appli
44	1759	85.7	3011	12	US-10-189-359-14	Sequence 14, Appli
45	1759	85.7	3011	14	US-10-259-275-20	Sequence 20, Appli

ALIGNMENTS

RESULT 1									
US-10-017-736-11									
1	Sequence 11, Application US/10017736								
2	Publication No. US2002019240A1								
3	GENERAL INFORMATION:								
4	APPLICANT: Boehringer Ingelheim (Canada) Ltd.								
5	FILE OF INVENTION: Purified Active HCV NS2/3 Protease								
6	FILE REFERENCE: 13/082								
7	CURRENT APPLICATION NUMBER: US/10/017,736								
8	PRIOR FILING DATE: 2001-12-14								
9	PRIOR APPLICATION NUMBER: 60/256,031								
10	NUMBER OF SEQ ID NOS: 21								
11	SOFTWARE: FastSeq for Windows Version 4.0								
12	SEQ ID NO 11								
13	LENGTH: 393								
14	TYPE: PRT								
15	ORGANISM: HCV								
16	US-10-017-736-11								
Query Match 100.0%; Score 2053; DB 13; Length 393;									
Best Local Similarity 100.0%; Pred. No. 2.2e-197;									
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MAASCGAVFGLALTLTSPYKVLARLIMWLYLIRVREAHLOWIPPLNVRGRDAI	60						
DB	1	MAASCGAVFGLALTLTSPYKVLARLIMWLYLIRVREAHLOWIPPLNVRGRDAI	60						
QY	61	ILITCAVHELIFDITKLLAIFGPMVLQAGITKPYFVRAGLIRACMVRKAAGHY	120						
DB	61	ILITCAVHELIFDITKLLAIFGPMVLQAGITKPYFVRAGLIRACMVRKAAGHY	120						
QY	121	VQMAFKALALGTYYDHLTFLQDMAHAGLFDLVANVPEVFSMEVEYLIIRWGADTAC	180						
DB	121	VQMAFKALALGTYYDHLTFLQDMAHAGLFDLVANVPEVFSMEVEYLIIRWGADTAC	180						
QY	181	GDIIISGLPYASARREIILGPDNFEQGMRLIAPITAYSQGTGILGCIITSLTGRDXN	240						

Db 181 GDIISGLPVASARRREIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDKN 240  
Qy 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Qy 301 PPGARSWTPTCCSSDLVYTRHADVI PVRRGDSKSLSPRPVSYLKSSGGPILCPG 360  
Db 301 PPGARSWTPTCCSSDLVYTRHADVI PVRRGDSKSLSPRPVSYLKSSGGPILCPG 360  
Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

## RESULT 2

US-10-650-585-11  
; Sequence 11, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; FILE OF INVENTION: Purified Active HCV NS2/3 Protease  
; TITLE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 393  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-650-585-11

Query Match 100.0%; Score 2053; DB 16; Length 393;  
Best Local Similarity 100.0%; Pred. No. 2, 2e-197;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHLOVWIPPLNVRGGDAI 60  
Db 1 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHLOVWIPPLNVRGGDAI 60  
Qy 61 ILITCAVHPELIFDITKLLAIFGFLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
Db 61 ILITCAVHPELIFDITKLLAIFGFLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
Qy 121 VQMAFMKLAALTGTYVDHLTPLODWAHAGLRDLAAVEPVISDMEXKIITWGAJTAAC 180  
Db 121 VQMAFMKLAALTGTYVDHLTPLODWAHAGLRDLAAVEPVISDMEXKIITWGAJTAAC 180  
Qy 181 GDIISGLPVASARRREIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDKN 240  
Db 181 GDIISGLPVASARRREIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDKN 240  
Qy 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLGPKPIOMYTNVDDLVGMQA 300  
Db 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLGPKPIOMYTNVDDLVGMQA 300  
Qy 301 PPGARSWTPTCCSSDLVYTRHADVI PVRRGDSKSLSPRPVSYLKSSGGPILCPG 360  
Db 301 PPGARSWTPTCCSSDLVYTRHADVI PVRRGDSKSLSPRPVSYLKSSGGPILCPG 360  
Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393

## RESULT 3

US-10-017-736-2

; Sequence 2, Application US/10017736

; Publication No. US20020192640A1

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; FILE OF INVENTION: Purified Active HCV NS2/3 Protease

; TITLE REFERENCE: 13/082

; CURRENT APPLICATION NUMBER: US/10/017,736

; CURRENT FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 409

; TYPE: PRT

; ORGANISM: HCV

US-10-017-736-2

Query Match 100.0%; Score 2053; DB 13; Length 409;  
Best Local Similarity 100.0%; Pred. No. 2, 3e-197;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHLOVWIPPLNVRGGDAI 60  
Db 5 MAASCGAVFGLALTLSPYKVLRLIMWLQYLITRVEAHLOVWIPPLNVRGGDAI 64  
Qy 61 ILITCAVHPELIFDITKLLAIFGFLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
Db 65 ILITCAVHPELIFDITKLLAIFGFLMVLOAGITKVPYFVRAQGLIRACMLVRKAAGHY 124  
Qy 121 VQMAFMKLAALTGTYVDHLTPLODWAHAGLRDLAAVEPVISDMEXKIITWGAJTAAC 180  
Db 125 VQMAFMKLAALTGTYVDHLTPLODWAHAGLRDLAAVEPVISDMEXKIITWGAJTAAC 184  
Qy 181 GDIISGLPVASARRREIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDKN 240  
Db 185 GDIISGLPVASARRREIILGPADNFEQGWRLAPITAYSQTRGLGCIITSLTGRDKN 244  
Qy 241 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 300  
Db 245 QVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAGKGPITOMYTNVDDLVGMQA 304  
Qy 301 PPGARSWTPTCCSSDLVYTRHADVI PVRRGDSKSLSPRPVSYLKSSGGPILCPG 360  
Db 305 PPGARSWTPTCCSSDLVYTRHADVI PVRRGDSKSLSPRPVSYLKSSGGPILCPG 364  
Qy 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393  
Db 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397

## RESULT 4

US-10-650-585-2

; Sequence 2, Application US/10650585

; Publication No. US20040077066A1

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; FILE OF INVENTION: Purified Active HCV NS2/3 Protease

; TITLE REFERENCE: 13/082

; CURRENT APPLICATION NUMBER: US/10/650,585

; CURRENT FILING DATE: 2003-08-28

; PRIOR APPLICATION NUMBER: US/10/017,736A

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 409

; TYPE: PRT

; ORGANISM: HCV

US-10-650-585-2

Query Match 100.0%; Score 2053; DB 16; Length 409;

Best Local Similarity 100.0%; Pred. No. 2,3e-197; Indels 0; Gaps 0;  
Matches 393; Conservative 0; Mismatches 0;

QY 1 MAASCGAVFIGIALITLSPYKVLARLIWMLOYLITREVAHLQWIPPLNVRGRDAI 60  
DB 5 MAASCGAVFIGIALITLSPYKVLARLIWMLOYLITREVAHLQWIPPLNVRGRDAI 64  
QY 61 ILITCAVHPELIFDITKLALIFGRLMVLQAGITKPYFVRAGGIRACMLVRKAAGHY 120  
DB 65 ILITCAVHPELIFDITKLALIFGRLMVLQAGITKPYFVRAGGIRACMLVRKAAGHY 124  
QY 121 VQVAFMKLAALITGYVDHLTPLODMAHAGRLDLAFAVEPVIFSDMEVKIITWGADTAAC 180  
DB 125 VQVAFMKLAALITGYVDHLTPLODMAHAGRLDLAFAVEPVIFSDMEVKIITWGADTAAC 184  
QY 181 GDIISGLPVSARRGRIELGPDNFGQWRLLAPITAYSOQTRGLGCIITSLGRDKN 240  
DB 185 GDIISGLPVSARRGRIELGPDNFGQWRLLAPITAYSOQTRGLGCIITSLGRDKN 244  
QY 241 QVEGEVQVSTATQSLFATCVNGVCTVPHGASGKTLAAGPKGPIITQMTNVODLVGMQA 300  
DB 245 QVEGEVQVSTATQSLFATCVNGVCTVPHGASGKTLAAGPKGPIITQMTNVODLVGMQA 304  
QY 301 PGARSMTECTCGSSDLVYVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPG 360  
DB 305 PGARSMTECTCGSSDLVYVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPG 364  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPEVESMETMR 393  
DB 365 GHAVGIFRAAVCTRGVAKAVDFIPEVESMETMR 397

RESULT 5  
US-10-017-736-12

; Sequence 12, Application US/10017736  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-017-736-12

Query Match 96.8%; Score 1987; DB 13; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9e-191;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ALLTISPYKVLARLIWMLOYLITREVAHLQWIPPLNVRGRDAIILITCAVHPELIF 73  
DB 1 ALLTISPYKVLARLIWMLOYLITREVAHLQWIPPLNVRGRDAIILITCAVHPELIF 60  
QY 74 DITKLALIFGRLMVLQAGITKPYFVRAGGIRACMLVRKAAGHYVQMAFMKLAALTG 133  
DB 61 DITKLALIFGRLMVLQAGITKPYFVRAGGIRACMLVRKAAGHYVQMAFMKLAALTG 120  
QY 134 TYVDHLTPLODMAHAGRLDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193  
DB 121 TYVDHLTPLODMAHAGRLDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
QY 194 GREILGPDNFGQWRLLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTAT 253  
DB 181 GREILGPDNFGQWRLLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTAT 240  
QY 254 QSFPLATCVNGVCTVPHGASGKTLAAGPKGPIITQMTNVODLVGMQAPGARSMTECTCG 313

DB 241 QSFPLATCVNGVCTVPHGASGKTLAAGPKGPIITQMTNVODLVGMQAPGARSMTECTCG 300

QY 314 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCT 373  
DB 301 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCT 360  
QY 374 RGVAKAVDFIPEVESMETMR 393  
DB 361 RGVAKAVDFIPEVESMETMR 380

RESULT 6  
US-10-650-585-12

; Sequence 12, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; PRIOR FILING DATE: 2003-08-28  
; CURRENT APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-12

Query Match 96.8%; Score 1987; DB 16; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9e-191;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ALLTISPYKVLARLIWMLOYLITREVAHLQWIPPLNVRGRDAIILITCAVHPELIF 73  
DB 1 ALLTISPYKVLARLIWMLOYLITREVAHLQWIPPLNVRGRDAIILITCAVHPELIF 60  
QY 74 DITKLALIFGRLMVLQAGITKPYFVRAGGIRACMLVRKAAGHYVQMAFMKLAALTG 133  
DB 61 DITKLALIFGRLMVLQAGITKPYFVRAGGIRACMLVRKAAGHYVQMAFMKLAALTG 120  
QY 134 TYVDHLTPLODMAHAGRLDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193  
DB 121 TYVDHLTPLODMAHAGRLDLAFAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
QY 194 GREILGPDNFGQWRLLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTAT 253  
DB 181 GREILGPDNFGQWRLLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTAT 240  
QY 254 QSFPLATCVNGVCTVPHGASGKTLAAGPKGPIITQMTNVODLVGMQAPGARSMTECTCG 313  
DB 241 QSFPLATCVNGVCTVPHGASGKTLAAGPKGPIITQMTNVODLVGMQAPGARSMTECTCG 300  
QY 314 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCT 373  
DB 301 SSDLYLVTRHADVIYVRRGDSRGLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCT 360  
QY 374 RGVAKAVDFIPEVESMETMR 393  
DB 361 RGVAKAVDFIPEVESMETMR 380

RESULT 7

US-10-029-907-3  
; Sequence 3, Application US/10029907  
; Publication No. US20020142350A1  
; GENERAL INFORMATION:  
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.

```

; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; TITLE OF INVENTION: HEPATITIS C VIRUS
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
; US-10-029-907-3

```

```

Query Match          95.0%; Score 1951; DB 13; Length 2201;
Best Local Similarity 93.1%; Pred. No. 4,4e-186;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 1 MAASCGAVFTGLALITLSPYKVLARLIMLYITRVEAHLOVWIPPLNVRGSDAI 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5 MAASCGAVFTGLALITLSPYKVLARLIMLYITRVEAHLOVWIPPLNVRGSDAI 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 ILTCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 ILTCAHPELIFITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 VQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAVEVIFSDMEVKIITWGADTAAC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 VQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAVEVIFSDMEVKIITWGADTAAC 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GDIIISGLPVSARRGREIHLGPADNLEGQWRLLAPITAYSOQTRGLGCIITSLGRDN 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 GDIIISGLPVSARRGREIHLGPADNLEGQWRLLAPITAYSOQTRGLGCIITSLGRDN 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAGKSGITQWYTNVDDLVGMQA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAGKSGITQWYTNVDDLVGMQA 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 PPGARSMPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 PPGARSMPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-10-309-561-3
; Sequence 3, Application US/10309561
; Publication No. US20030148348A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US/10/029,907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3

```

```

; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
; US-10-309-561-3

```

```

Query Match          95.0%; Score 1951; DB 14; Length 2201;
Best Local Similarity 93.1%; Pred. No. 4,4e-186;
Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 1 MAASCGAVFTGLALITLSPYKVLARLIMLYITRVEAHLOVWIPPLNVRGSDAI 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5 MAASCGAVFTGLALITLSPYKVLARLIMLYITRVEAHLOVWIPPLNVRGSDAI 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 ILTCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 ILTCAHPELIFITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 VQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAVEVIFSDMEVKIITWGADTAAC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 VQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAVEVIFSDMEVKIITWGADTAAC 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GDIIISGLPVSARRGREIHLGPADNLEGQWRLLAPITAYSOQTRGLGCIITSLGRDN 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 GDIIISGLPVSARRGREIHLGPADNLEGQWRLLAPITAYSOQTRGLGCIITSLGRDN 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAGKSGITQWYTNVDDLVGMQA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 QVEGEVQVSTATOSFLATCVNGVCWTVFHGAGSKTLAGKSGITQWYTNVDDLVGMQA 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 PPGARSMPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 PPGARSMPTCTCGSSDLYLTRHADVIYVRRRGDSRSLSPRPVSYLKSSGGPILCP 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 GHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
US-10-467-000-1
; Sequence 1, Application US/10467000
; Publication No. US20040067486A1
; GENERAL INFORMATION:
; APPLICANT: De Francesco, Raffaele
; APPLICANT: Migliaccio, Giovanni
; APPLICANT: Paolesse, Giacomo
; TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
; FILE REFERENCE: ITR0003P
; CURRENT APPLICATION NUMBER: US/10/467,000
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: PCT/EP02/00526
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/263,479
; PRIOR FILING DATE: 2001-01-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Con 1 HCV isolate nucleic acid
; US-10-467-000-1

```

```

Query Match          95.0%; Score 1951; DB 12; Length 3010;
Best Local Similarity 93.1%; Pred. No. 6,9e-186;

```

Matches 366; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALITLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60  
DB 814 MAASCGAVFGLITLSPHYKLFLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAV 873

QY 61 ILITCAVPELITFDITKLILAIIFGPIMLVQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
DB 874 ILITCAVPELITFDITKLILAIIFGPIMLVQAGITKVPYFVRAQGLIRACMLVRKAAGHY 933

QY 121 VQMAFMKLAULTGTYYVDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 180  
DB 934 VQMAFMKLAULTGTYYVDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 993

QY 181 GDIISGLPVSARRGREIILGPADNFEQGWRLLAPITAYSQOTRGLGCIITSLTGRDN 240  
DB 994 GDIISGLPVSARRGREIILGPADNFEQGWRLLAPITAYSQOTRGLGCIITSLTGRDN 1053

QY 241 QVEGEVQVSTATQSFATCNGVCMTVPHGAGSKTLAGPKPITOMTYNVDDLVGMQA 300  
DB 1054 QVEGEVQVSTATQSFATCNGVCMTVPHGAGSKTLAGPKPITOMTYNVDDLVGMQA 1113

QY 301 PPGASMTPTCTGSSDLVLTVRHADVIIVRRRDSRGSLSLSPRVSYLKSGSGGPLLCP 360  
DB 1114 PPGASMTPTCTGSSDLVLTVRHADVIIVRRRDSRGSLSLSPRVSYLKSGSGGPLLCP 1173

QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 1174 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 10  
US-09-919-901-4  
Sequence 4, Application US/09919901  
Publication No. US2003082518A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-4

Query Match 94.8%; Score 1946; DB 10; Length 1692;  
Best Local Similarity 92.9%; Pred. No. 9,7e-186;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 181 GDIISGLPVSARRGREIILGPADNFEQGWRLLAPITAYSQOTRGLGCIITSLTGRDN 240  
DB 273 GDIISGLPVSARRGREIILGPADNFEQGWRLLAPITAYSQOTRGLGCIITSLTGRDN 332

QY 241 QVEGEVQVSTATQSFATCNGVCMTVPHGAGSKTLAGPKPITOMTYNVDDLVGMQA 300  
DB 333 QVEGEVQVSTATQSFATCNGVCMTVPHGAGSKTLAGPKPITOMTYNVDDLVGMQA 392

QY 301 PPGASMTPTCTGSSDLVLTVRHADVIIVRRRDSRGSLSLSPRVSYLKSGSGGPLLCP 360  
DB 393 PPGASMTPTCTGSSDLVLTVRHADVIIVRRRDSRGSLSLSPRVSYLKSGSGGPLLCP 452

QY 361 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 485

RESULT 11  
US-10-191-966-4  
Sequence 4, Application US/10191966  
Publication No. US20030175692A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/10/191,966  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: US/09/263,933  
PRIOR FILING DATE: 1999-03-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-10-191-966-4

Query Match 94.8%; Score 1946; DB 14; Length 1692;  
Best Local Similarity 92.9%; Pred. No. 9,7e-186;  
Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALITLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60  
DB 93 MAASCGAVFGLVLTLSPPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 152

QY 61 ILITCAVPELITFDITKLILAIIFGPIMLVQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
DB 153 ILITCAVPELITFDITKLILAIIFGPIMLVQAGITKVPYFVRAQGLIRACMLVRKAAGHY 212

QY 121 VQMAFMKLAULTGTYYVDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 180  
DB 213 VQMAFMKLAULTGTYYVDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAAC 272

QY 181 GDIISGLPVSARRGREIILGPADNFEQGWRLLAPITAYSQOTRGLGCIITSLTGRDN 240  
DB 273 GDIISGLPVSARRGREIILGPADNFEQGWRLLAPITAYSQOTRGLGCIITSLTGRDN 332

QY 241 QVEGEVQVSTATQSFATCNGVCMTVPHGAGSKTLAGPKPITOMTYNVDDLVGMQA 300  
DB 333 QVEGEVQVSTATQSFATCNGVCMTVPHGAGSKTLAGPKPITOMTYNVDDLVGMQA 392

QY 301 PPGASMTPTCTGSSDLVLTVRHADVIIVRRRDSRGSLSLSPRVSYLKSGSGGPLLCP 360  
DB 393 PPGASMTPTCTGSSDLVLTVRHADVIIVRRRDSRGSLSLSPRVSYLKSGSGGPLLCP 452

QY 361 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
 DB 453 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 485

## RESULT 12

US-09-919-901-2  
 ; Sequence 2, Application US/09919901  
 ; Publication No. US2003082518A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263,933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2307  
 ; TYPE: PRF  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 US-09-919-901-2

Query Match 94.8%; Score 1946; DB 10; Length 2307;  
 Best Local Similarity 92.9%; Pred. No. 1.5e-185;

Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60  
 DB 185 MAASCGAVFGLVLLTSPYKVELARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 244  
 QY 61 ILTCAVHPELIPITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 245 ILMCAVHPELIPITKLLAIFGLPLVLOAGITRVPYFRAQGLIRACMLVRKAGGHY 304  
 QY 121 VQMAFMKLAALTGYYVDHLTPLODMAHAGLRDLAVALVEPVISDMVEKLIITGADTAAC 180  
 DB 305 VQMAFMKLGALTGYIYNHLTPLODMAHAGLRDLAVALVEPVISDMETKIITGADTAAC 364  
 QY 181 GDITSGLPVSARRREILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 240  
 DB 365 GDITSGLPVSARRREILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 424  
 QY 241 QVEGEVQVSTATQSFATCNGVCMVTFPGAGSKTLAAGKPIITOMYTNVQDVLVQMA 300  
 DB 425 QVEGEVQVSTATQSFATCNGVCMVTFPGAGSKTLAAGKPIITOMYTNVQDVLVQMA 484  
 QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPVRRRSGSRGSLSPRVSYLKSSGGPLICPS 360  
 DB 485 PGARSMTPTCTGSSDLYLVTRHADVIPVRRRSGSRGSLSPRVSYLKSSGGPLICPS 544  
 QY 361 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
 DB 545 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 577

## RESULT 13

US-10-191-966-2  
 ; Sequence 2, Application US/10191966  
 ; Publication No. US20030175692A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.

; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/10/191,966  
 ; CURRENT FILING DATE: 2002-07-10  
 ; PRIOR APPLICATION NUMBER: US/09/263,933  
 ; PRIOR FILING DATE: 1999-03-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2307  
 ; TYPE: PRF  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 US-10-191-966-2

Query Match 94.8%; Score 1946; DB 14; Length 2307;  
 Best Local Similarity 92.9%; Pred. No. 1.5e-185;  
 Matches 365; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFGLALTLSPYKVLARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 60  
 DB 185 MAASCGAVFGLVLLTSPYKVELARLIWLOYLITRVEAHLQWIPPLNVRGGRDAI 244  
 QY 61 ILTCAVHPELIPITKLLAIFGLPLVLOAGITKVPYFRAQGLIRACMLVRKAGGHY 120  
 DB 245 ILMCAVHPELIPITKLLAIFGLPLVLOAGITRVPYFRAQGLIRACMLVRKAGGHY 304  
 QY 121 VQMAFMKLAALTGYYVDHLTPLODMAHAGLRDLAVALVEPVISDMVEKLIITGADTAAC 180  
 DB 305 VQMAFMKLGALTGYIYNHLTPLODMAHAGLRDLAVALVEPVISDMETKIITGADTAAC 364  
 QY 181 GDITSGLPVSARRREILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 240  
 DB 365 GDITSGLPVSARRREILLGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKN 424  
 QY 241 QVEGEVQVSTATQSFATCNGVCMVTFPGAGSKTLAAGKPIITOMYTNVQDVLVQMA 300  
 DB 425 QVEGEVQVSTATQSFATCNGVCMVTFPGAGSKTLAAGKPIITOMYTNVQDVLVQMA 484  
 QY 301 PGARSMTPTCTGSSDLYLVTRHADVIPVRRRSGSRGSLSPRVSYLKSSGGPLICPS 360  
 DB 485 PGARSMTPTCTGSSDLYLVTRHADVIPVRRRSGSRGSLSPRVSYLKSSGGPLICPS 544  
 QY 361 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 393  
 DB 545 GHAAGIFRAAVCTRGVAKAVDFPVESMETTMR 577

## RESULT 14

US-09-919-901-11  
 ; Sequence 11, Application US/09919901  
 ; Publication No. US2003082518A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263,933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 1692



TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-11

Query Match 94.6%; Score 1943; DB 10; Length 1692;  
Best Local Similarity 92.6%; Pred. No. 1.9e-185;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGANFIGIALITLSPYKVLARLIWLOQITITVEAHLQVWIPPLNVRGGRDAI 60  
DB 93 MAASCGAVFVGIVLTLSPYKVFARLIWLOQITITVEAHLQVWIPPLNVRGGRDAI 152  
QY 61 ILTCAVHPELIDITKLLAIIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
DB 153 ILLMCVHPELIDITKLLAIIFGRLMVLQAGITRVPYFVRAQGLIRACMLVRKAAGHY 212  
QY 121 VQMAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAVEVIFSDMEVKIITWGADTAAC 180  
DB 213 VQMAFMKLAGLTGYIYNHLTPLRDMAHAGLRDLAVAVEVIFSDMETKIITWGADTAAC 272  
QY 181 GDITSLPVSARRGREILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDKN 240  
DB 273 GDITSLPVSARRGREILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDKN 332  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDDLVGMQA 300  
DB 333 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDDLVGMQA 392  
QY 301 PPGARSMTPCTCGSSDLVYVTRHADVIIVRRRGDSRGLSPREVSYLKSGSGGFLCPS 360  
DB 393 PPGARSLTPCTCGSSDLVYVTRHADVIIVRRRGDSRGLSPREVSYLKSGAGGFLCPS 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 485

RESULT 15  
US-10-191-966-11  
; Sequence 11, Application US/10191966  
; Publication No. US20030175692A1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/10/191,966  
; CURRENT FILING DATE: 2002-07-10  
; PRIOR APPLICATION NUMBER: US/09/263,933  
; PRIOR FILING DATE: 1999-03-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: :  
US-10-191-966-11

Query Match 94.6%; Score 1943; DB 14; Length 1692;  
Best Local Similarity 92.6%; Pred. No. 1.9e-185;  
Matches 364; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAASCGAVFVGLTLSPYKVLARLIWLOQITITVEAHLQVWIPPLNVRGGRDAI 60  
DB 93 MAASCGAVFVGLTLSPYKVLARLIWLOQITITVEAHLQVWIPPLNVRGGRDAI 152

QY 61 ILTCAVHPELIDITKLLAIIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHY 120  
DB 153 ILLMCVHPELIDITKLLAIIFGRLMVLQAGITRVPYFVRAQGLIRACMLVRKAAGHY 212  
QY 121 VQMAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAVEVIFSDMEVKIITWGADTAAC 180  
DB 213 VQMAFMKLAGLTGYIYNHLTPLRDMAHAGLRDLAVAVEVIFSDMETKIITWGADTAAC 272  
QY 181 GDITSLPVSARRGREILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDKN 240  
DB 273 GDITSLPVSARRGREILGPADNFEQGRLLAPITAYSQOTRGLIGCIITSLTGRDKN 332  
QY 241 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDDLVGMQA 300  
DB 333 QVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDDLVGMQA 392  
QY 301 PPGARSMTPCTCGSSDLVYVTRHADVIIVRRRGDSRGLSPREVSYLKSGSGGFLCPS 360  
DB 393 PPGARSLTPCTCGSSDLVYVTRHADVIIVRRRGDSRGLSPREVSYLKSGAGGFLCPS 452  
QY 361 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 393  
DB 453 GHAVGIFRAAVCTRGVAKAVDFIVESMETTMR 485

Search completed: May 6, 2004, 09:43:19  
Job time : 41.2025 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 51.7244 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-12  
Sequence: 1 ALLTSPYKYLARLIMWL.....RGVAKAVDFPVESEMTNR 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*\*  
1: geneseqp1980s:\*\*  
2: geneseqp1990s:\*\*  
3: geneseqp2000s:\*\*  
4: geneseqp2001s:\*\*  
5: geneseqp2002s:\*\*  
6: geneseqp2003as:\*\*  
7: geneseqp2003bs:\*\*  
8: geneseqp2004s:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1987	100.0	380	5	ABG32185 HCV prote
2	1987	100.0	393	5	ABG32184 HCV prote
3	1987	100.0	409	5	ABG32181 HCV prote
4	1902	95.7	3010	2	AAR87694 Partial H
5	1897	95.5	3010	2	AAR86822 HCV prote
6	1896	95.4	3010	2	AAR86864 Hepatitis
7	1888	95.0	768	2	AAR40223 Recombina
8	1887	95.0	2201	5	ABG30601 Hepatitis
9	1887	95.0	2201	5	ABG30591 Hepatitis
10	1887	95.0	2201	5	ABG30600 Hepatitis
11	1887	95.0	2201	5	ABG30581 Hepatitis
12	1887	95.0	2201	5	ABG30593 Hepatitis
13	1887	95.0	2201	5	ABG30582 Hepatitis
14	1887	95.0	2201	5	ABG30580 Hepatitis
15	1887	95.0	2201	5	ABG30587 Hepatitis
16	1887	95.0	2201	5	ABG30599 Hepatitis
17	1887	95.0	2201	5	ABG30594 Hepatitis
18	1887	95.0	2201	5	ABG30598 Hepatitis
19	1887	95.0	2201	5	ABG30595 Hepatitis
20	1887	95.0	3010	5	ABG32458 Hepatitis
21	1887	95.0	3010	5	ABG32459 Hepatitis
22	1887	95.0	3010	5	ABG32451 Hepatitis
23	1887	95.0	3010	5	ABG32455 Hepatitis
24	1887	95.0	3010	5	ABG32457 Hepatitis
25	1887	95.0	3010	5	ABG32460 Hepatitis

26	1887	95.0	3010	5	ABG32461 Hepatitis
27	1887	95.0	3010	5	ABG32454 Hepatitis
28	1887	95.0	3011	5	ABG32456 Hepatitis
29	1884	94.8	2201	5	ABG30586 Hepatitis
30	1884	94.8	2201	5	ABG30583 Hepatitis
31	1884	94.8	2201	5	ABG30589 Hepatitis
32	1884	94.8	2201	5	ABG30588 Hepatitis
33	1883	94.8	2201	5	ABG30590 Hepatitis
34	1881	94.7	2307	3	ABG32452 Hepatitis
35	1881	94.7	3010	5	ABG30584 Hepatitis
36	1880	94.6	2201	5	ABG30582 Hepatitis
37	1880	94.6	2201	5	ABG30602 Hepatitis
38	1880	94.6	3010	5	ABG32453 Hepatitis
39	1878	94.5	2307	3	AAV70065 Recombina
40	1878	94.5	3014	2	AAV54099 NANBHV E1
41	1876	94.5	2201	5	ABG30585 Hepatitis
42	1873	94.3	3014	2	AAV35207 Hepatitis
43	1872	94.2	3090	7	ADD67962 EMCV inte
44	1869	94.1	2307	3	AAV70066 Recombina
45	1869	94.1	3010	2	AAW98022 Infection

## ALIGNMENTS

RESULT 1  
ID ABG32185 standard; proteain. 380 AA.  
XX  
AC ABG32185;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 827-1206.  
XX  
KW HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
KW chronic liver disease; cirrhosis; end-stage liver disease; virologic;  
KW hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KW chaotropic agent; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200248375-A2.  
XX  
PD 20-JUN-2002.  
XX  
PF 13-DEC-2001; 2001MO-CA001796.  
XX  
PR 15-DEC-2000; 2000US-0256031P.  
XX  
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
PI Thibault D, Lamare D, Maurice R, Pilote L, Pause A;  
XX WPI, 2002-599511/64.  
XX  
DR Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
PS Claim 41; Page 60-61; 67bp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32184, (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 380 AA;

Query Match 100.0%; Score 1987; DB 5; Length 380;

Best Local Similarity 100.0%; Pred. No. 1.5e-184; Indels 0; Gaps 0;

Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIMWLYITVEAHLQWIPPLNVRGGRDAIILLTCVHPELIF 60  
 Db 1 ALLTSPYKVLARLIMWLYITVEAHLQWIPPLNVRGGRDAIILLTCVHPELIF 60  
 QY 61 DITKLLAIFGPIMLVQAGITKPYFVRAGQIRACMLVRKKAAGHYVOMAFKLAALTG 120  
 Db 61 DITKLLAIFGPIMLVQAGITKPYFVRAGQIRACMLVRKKAAGHYVOMAFKLAALTG 120  
 QY 121 TYYVDHLTFLQDMAHAGLNDLAAVEPVFSDMEVKIITWGADTACGDIISGLPVSARR 180  
 Db 121 TYYVDHLTFLQDMAHAGLNDLAAVEPVFSDMEVKIITWGADTACGDIISGLPVSARR 180  
 QY 121 TYYVDHLTFLQDMAHAGLNDLAAVEPVFSDMEVKIITWGADTACGDIISGLPVSARR 180  
 Db 121 TYYVDHLTFLQDMAHAGLNDLAAVEPVFSDMEVKIITWGADTACGDIISGLPVSARR 180  
 QY 181 GREILLGPADNFEQGQWRLLAPITAYVSGQTRGLGCIITSLTGRDNQYEGEVQVSTAT 240  
 Db 181 GREILLGPADNFEQGQWRLLAPITAYVSGQTRGLGCIITSLTGRDNQYEGEVQVSTAT 240  
 QY 241 QSEFLATCVAGVCMVTFHGAGSKTLAAGPKPITOMTNNVDLVGMQAPRGASMTPTCG 300  
 Db 241 QSEFLATCVAGVCMVTFHGAGSKTLAAGPKPITOMTNNVDLVGMQAPRGASMTPTCG 300  
 QY 301 SSDLYLVTRHADVIPIVRRGRDGRGSLSPRVSYLKSSGGPILLCPSGHAGVIFRAAVCT 360  
 Db 301 SSDLYLVTRHADVIPIVRRGRDGRGSLSPRVSYLKSSGGPILLCPSGHAGVIFRAAVCT 360  
 QY 361 RGAKAVDPIPVESMETTWK 380  
 Db 361 RGAKAVDPIPVESMETTWK 380

RESULT 2

ABG32184  
 ID ABG32184 standard; protein; 393 AA.

XX ABG32184;

XX 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 815-1206.

KM HCV, enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viricide;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.

XX Hepatitis C virus.  
 OS Synthetic.  
 XX MO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-CAN01796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 41; Page 59-60; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32184; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 393 AA;

Query Match 100.0%; Score 1987; DB 5; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.6e-184; Indels 0; Gaps 0;

Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIMWLYITVEAHLQWIPPLNVRGGRDAIILLTCVHPELIF 60  
 Db 1 ALLTSPYKVLARLIMWLYITVEAHLQWIPPLNVRGGRDAIILLTCVHPELIF 60  
 QY 14 ALLTSPYKVLARLIMWLYITVEAHLQWIPPLNVRGGRDAIILLTCVHPELIF 73  
 Db 14 ALLTSPYKVLARLIMWLYITVEAHLQWIPPLNVRGGRDAIILLTCVHPELIF 73  
 QY 61 DITKLLAIFGPIMLVQAGITKPYFVRAGQIRACMLVRKKAAGHYVOMAFKLAALTG 120  
 Db 61 DITKLLAIFGPIMLVQAGITKPYFVRAGQIRACMLVRKKAAGHYVOMAFKLAALTG 120  
 QY 74 DITKLLAIFGPIMLVQAGITKPYFVRAGQIRACMLVRKKAAGHYVOMAFKLAALTG 133  
 Db 74 DITKLLAIFGPIMLVQAGITKPYFVRAGQIRACMLVRKKAAGHYVOMAFKLAALTG 133  
 QY 121 TYYVDHLTFLQDMAHAGLNDLAAVEPVFSDMEVKIITWGADTACGDIISGLPVSARR 180  
 Db 121 TYYVDHLTFLQDMAHAGLNDLAAVEPVFSDMEVKIITWGADTACGDIISGLPVSARR 180

DB 134 TVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGDADPAACDIIISGLPVSARR 193  
QY 181 GREILLGPADNFEQGGWRLAPITAVSQOTRGLICITSLTGRDKNOVEGEVQVSTAT 240  
DB 194 GREILLGPADNFEQGGWRLAPITAVSQOTRGLICITSLTGRDKNOVEGEVQVSTAT 253  
QY 241 OSFLATCVNGVCMVTFHAGSGKTLAAGPKPITQMTNTNDODLVGMQAPPGARSMTPTCTG 300  
DB 254 OSFLATCVNGVCMVTFHAGSGKTLAAGPKPITQMTNTNDODLVGMQAPPGARSMTPTCTG 313  
QY 301 SSDLVLTVRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLLCPSGHAGVIFPAAVCT 360  
DB 314 SSDLVLTVRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLLCPSGHAGVIFPAAVCT 373  
QY 361 RGVAKAVDFIPVESMETTMR 380  
DB 374 RGVAKAVDFIPVESMETTMR 393

RESULT 3  
ABG32181  
ID ABG32181 standard; protein; 409 AA.

ABG32181;

05-NOV-2002 (first entry)

HCV protease NS2/3 (810-1206).

HCV: enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
hepatotropic; antiinflammatory; lauryldiethyamine oxide; LDAO;  
chaotropic agent; mutant; mutagen.

Hepatitis C virus.  
Synthetic.

Key Location/Qualifiers  
Peptide 398..409  
/note="Streptavidin tag"

W0200248375-A2.

20-JUN-2002.

13-DEC-2001; 2001MO-CA001796.

15-DEC-2000; 2000US-0256031P.

(BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

WPI, 2002-559511/64.  
N-PSDB; ABK90406.

Novel polypeptide for screening inhibitors of non-structural proteases  
useful as therapeutic agents against hepatitis C virus, comprises full  
length non-structural protease, or its truncation.

Claim 42; Fig 1B; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
residue amino acid 810 to 906, or having a minimal amino acid sequence  
from residue 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
NS2/3 protease. Also included are (1) a composition (C) comprising an  
isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
its truncation or a mutated sequence, where the protease is in a solution  
comprising a sufficient concentration of lauryldiethyamine oxide (LDAO)  
to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 (810-1206) protein, which has a C-  
CC terminal streptavidin tag

SQ Sequence 409 AA;

Query Match 100.0%; Score 1987; DB 5; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1,6e-184;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKYLARLWLOYLITRYBAHQWITPELNVRGSDAIIILTCANPELIF 60  
DB 18 ALLTSPYKYLARLWLOYLITRYBAHQWITPELNVRGSDAIIILTCANPELIF 77  
QY 61 DITKLLAIFGLPLVLOAGITKVPYFAAGLIRACMIVRAAGHYQVAFMLAALTG 120  
DB 78 DITKLLAIFGLPLVLOAGITKVPYFAAGLIRACMIVRAAGHYQVAFMLAALTG 137  
QY 121 TVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGDADPAACDIIISGLPVSARR 180  
DB 138 TVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGDADPAACDIIISGLPVSARR 197  
QY 181 GREILLGPADNFEQGGWRLAPITAVSQOTRGLICITSLTGRDKNOVEGEVQVSTAT 240  
DB 198 GREILLGPADNFEQGGWRLAPITAVSQOTRGLICITSLTGRDKNOVEGEVQVSTAT 257  
QY 241 OSFLATCVNGVCMVTFHAGSGKTLAAGPKPITQMTNTNDODLVGMQAPPGARSMTPTCTG 300  
DB 258 OSFLATCVNGVCMVTFHAGSGKTLAAGPKPITQMTNTNDODLVGMQAPPGARSMTPTCTG 317  
QY 301 SSDLVLTVRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLLCPSGHAGVIFPAAVCT 360  
DB 318 SSDLVLTVRHADVTPVRRRGDSRGLSPRPVSYLKSGSGGPLLCPSGHAGVIFPAAVCT 377  
QY 361 RGVAKAVDFIPVESMETTMR 380  
DB 378 RGVAKAVDFIPVESMETTMR 397

RESULT 4  
AAR82694  
ID AAR82694 standard; protein; 3010 AA.

AAR82694;

16-OCT-2003 (revised)  
DT 14-NOV-1996 (first entry)

Partial HCV non-structural polyprotein.

protease; hepatitis C virus; screening; inhibitor; proteolytic;  
identification; cleavage.

XX

OS Hepatitis C virus; Virus.  
 XX Key Location/Qualifiers  
 XX 898. .1233  
 FT Protein /note="Partial proteinase; see AAR82692"  
 FT Protein 992. .1907  
 FT /note="partial proteinase; see AAR82693"  
 XX JF07184648-A.  
 XX 25-JUL-1995.  
 XX PD  
 XX PF 05-FEB-1993; 93JP-00018854.  
 XX PR 07-FEB-1992; 92JP-00022657.  
 PR 18-SEP-1992; 92JP-00249240.  
 PR 04-DEC-1992; 92JP-00325303.  
 XX (KAEN/) KAENNO K.  
 PA (SUMO) SUMITOMO METAL IND LTD.  
 PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX WPI; 1995-287962/38.  
 DR N-PSDB; AAT03960.  
 XX An HCV proteinase active substance - which has activity as an anti-HCV  
 PT agent and can be used to screen for proteinase inhibitors.  
 XX Disclosure; Page 39-48; 52pp; Japanese.  
 CC The present sequence is a partial Hepatitis C virus (HCV) polyprotein  
 CC from the non-structural region. Partial proteinase sequences (AAR82692-  
 CC 93) are contained within this sequence. The proteinases can be used as  
 CC anti-HCV agents. They can also be used to screen cpgs. for their ability  
 CC to inhibit their proteolytic activity. In this way proteinase inhibitors  
 CC can be identified. (updated on 16-OCT-2003 to standardise OS field)  
 XX SQ Sequence 3010 AA;  
 Query Match 95.7%; Score 1902; DB 2; Length 3010;  
 Best Local Similarity 94.2%; Pred. No. 5.3e-175;  
 Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 LTLSPYKYLARLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61  
 DB 828 LTLSPYKYLARLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 887  
 QY 62 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFKLAALTGT 121  
 DB 888 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFKLAALTGT 947  
 QY 122 YVYDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIITWGAATAAGDIISGLPVSARG 181  
 DB 948 YVYDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIITWGAATAAGDIISGLPVSARG 1007  
 QY 182 REILGPRADNFGGQWRLLAPITAYSQOTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 241  
 DB 1008 KEILLGPADSFEGQWRLLAPITAYSQOTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 1067  
 QY 242 SFLATCVNGVCTVTHAGAGSKTLAGKPGITOMYTNVDDLVGMQAPPGASMTPTCTGS 301  
 DB 1068 SFLATCVNGVCTVTHAGAGSKTLAGKPGITOMYTNVDDLVGMQAPPGASMTPTCTGS 1127  
 QY 302 SDLYVTNRADYIPVRRGDSRGLSPRVSYLKSGSGGPRLLCPSGHVAIGIFRAAVCTR 361  
 DB 1128 SDLYVTNRADYIPVRRGDSRGLSPRVSYLKSGSGGPRLLCPSGHVAIGIFRAAVCTR 1187  
 QY 362 GVAKAVDFIPVESMETTMR 380  
 DB 1186 GVAKAVDFIPVESMETTMR 1206  
 RESULT 5

AAR86822  
 ID AAR86822 standard; protein; 3010 AA.  
 XX AAR86822;  
 AC AAR86822;  
 XX 16-OCT-2003 (revised)  
 DT 16-OCT-1995 (first entry)  
 DE HCV protein cleavable with new serine proteinase.  
 XX HCV proteinase; serine; cleavage; hepatitis C virus; HCV.  
 XX Hepatitis C virus; Virus.  
 XX Key Location/Qualifiers  
 FT Cleavage-site 2419. .2420  
 FT /note="Serine protease cleavage site"  
 XX JF06315377-A.  
 XX 15-NOV-1994.  
 XX PD  
 XX PF 06-MAY-1993; 93JP-00105666.  
 XX PR 06-MAY-1993; 93JP-00105666.  
 XX (KAEN/) KAENNO K.  
 PA (SUMO) SUMITOMO METAL IND LTD.  
 PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX WPI; 1995-032330/05.  
 DR N-PSDB; AAO80498.  
 XX New HCV-originated proteinase active substance - used for site-specific  
 PT cleavage by an intermolecular reaction and the purification thereof.  
 XX Disclosure; Page 10-19; 23pp; Japanese.  
 CC This protein from HCV (hepatitis C virus) (encoded by AAO80498) is  
 CC cleaved between amino acids 2419 and 2420, by a new serine protease.  
 CC contg. the sequence of AAR86821. The proteinase is purified as a fused  
 CC product with the dihydrofolate reductase protein by using a methotrexate  
 CC column. It can be used for the development of an inhibitor for HCV  
 CC proteinase. (updated on 16-OCT-2003 to standardise OS field)  
 XX SQ Sequence 3010 AA;  
 Query Match 95.5%; Score 1897; DB 2; Length 3010;  
 Best Local Similarity 93.9%; Pred. No. 1.6e-174;  
 Matches 356; Conservative 12; Mismatches 11; Indels 0; Gaps 0;  
 QY 2 LTLSPYKYLARLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61  
 DB 828 LTLSPYKYLARLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 887  
 QY 62 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFKLAALTGT 121  
 DB 888 ITKLLAIFGPIMLVQAGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFKLAALTGT 947  
 QY 122 YVYDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIITWGAATAAGDIISGLPVSARG 181  
 DB 948 YVYDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIITWGAATAAGDIISGLPVSARG 1007  
 QY 182 REILGPRADNFGGQWRLLAPITAYSQOTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 241  
 DB 1008 KEILLGPADSFEGQWRLLAPITAYSQOTRGLGCIITSLTRGRDNQVGEVQVNSTATQ 1067  
 QY 242 SFLATCVNGVCTVTHAGAGSKTLAGKPGITOMYTNVDDLVGMQAPPGASMTPTCTGS 301  
 DB 1068 SFLATCVNGVCTVTHAGAGSKTLAGKPGITOMYTNVDDLVGMQAPPGASMTPTCTGS 1127  
 QY 302 SDLYVTNRADYIPVRRGDSRGLSPRVSYLKSGSGGPRLLCPSGHVAIGIFRAAVCTR 361

DB 1128 SDLYLVTRHADVPVRRRGRSGSLSPRPISYLKSSGGPILCPGSHVVGIFRAAVCTR 1187  
 QY 362 GVAKAVDPIPVESMETTMR 380  
 DB 1188 GVAKAVDPIPVESMETTMR 1206

RESULT 6  
 AAR68864  
 ID AAR68864 standard; protein; 3010 AA.  
 XX AAR68864;  
 AC 06-DEC-1995 (first entry)  
 DT 06-DEC-1995 (first entry)  
 DE Hepatitis C virus RNA helicase.  
 XX Hepatitis C virus RNA helicase.  
 KM Hepatitis C virus; HCV, non-A non-B; helicase gene; RNA helicase;  
 KM Baculovirus; recombinant production.  
 XX Hepatitis C virus.  
 OS

Key Location/Qualifiers  
 FH 196..198  
 FT /label= N-linked glycosylation site  
 FT 209..211  
 FT /label= N-linked glycosylation site  
 FT 234..236  
 FT /label= N-linked glycosylation site  
 FT 250..252  
 FT /label= N-linked glycosylation site  
 FT 305..307  
 FT /label= N-linked glycosylation site  
 FT 325..327  
 FT /label= N-linked glycosylation site  
 FT 417..419  
 FT /label= N-linked glycosylation site  
 FT 423..425  
 FT /label= N-linked glycosylation site  
 FT 430..432  
 FT /label= N-linked glycosylation site  
 FT 448..450  
 FT /label= N-linked glycosylation site  
 FT 532..534  
 FT /label= N-linked glycosylation site  
 FT 556..558  
 FT /label= N-linked glycosylation site  
 FT 576..578  
 FT /label= N-linked glycosylation site  
 FT 623..625  
 FT /label= N-linked glycosylation site  
 FT 645..647  
 FT /label= N-linked glycosylation site  
 FT 1213..1215  
 FT /label= N-linked glycosylation site  
 FT 1255..1257  
 FT /label= N-linked glycosylation site  
 FT 2041..2043  
 FT /label= N-linked glycosylation site  
 FT 2077..2079  
 FT /label= N-linked glycosylation site  
 FT 2240..2242  
 FT /label= N-linked glycosylation site  
 FT 2788..2790  
 FT /label= N-linked glycosylation site  
 FT Region  
 XX JP06319583-A.  
 XX 22-NOV-1994.  
 XX 18-SEP-1992; 92JF-00249241.  
 XX 18-SEP-1992; 92JF-00249241.  
 PR

XX (SOYA-) SOYAKU GIUTSU KENKUTSUO KK.  
 PA WPI; 1995-040330/06.  
 DR N-PSDB; AA081559.  
 DR  
 XX  
 PT of hepatitis C virus helicase gene in baculovirus - useful for large  
 PT scale prodn. of RNA helicase.  
 PS Claim 1, Fig 1-4; 9pp; Japanese.  
 XX  
 CC AA081559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The DNA  
 CC was used in the construction of an expression vector, which was used to  
 CC transform a baculovirus host. The transformed baculovirus could then be  
 CC used for the recombinant prodn. of HCV RNA helicase  
 XX

Sequence 3010 AA;  
 SQ

Query Match 95.4%; Score 1896; DB 2; Length 3010;  
 Best Local Similarity 93.9%; Pred. No. 2e-174;  
 Matches 356; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 2 LITLSPYKVLARLIMWLOVLRVRAHLQWMIPIPNVNGSGDAILLTCVAHPHLIFD 61  
 DB LITLSPYKVLARLIMWLOVLRVRAHLQWMIPIPNVNGSGDAILLTCVAHPHLIFD 887  
 QY 62 ITKLLALFGEPLMTLQAGITKVPYFVRAQGLIRACMLVRRKAGSHYQNAFMKLAALTGT 121  
 DB ITKLLALFGEPLMTLQAGITKVPYFVRAQGLIRACMLVRRKAGSHYQNAFMKLAALTGT 947  
 QY 122 VYVDHLPLDQMAWAGRDIAVAVEPIPSDMETKLTWGAADTAAGDIISGLPVBARBG 181  
 DB VYVDHLPLDQMAWAGRDIAVAVEPIPSDMETKLTWGAADTAAGDIISGLPVBARBG 1007  
 QY 948 VYVDHLPLDQMAWAGRDIAVAVEPIPSDMETKLTWGAADTAAGDIISGLPVBARBG 1007  
 DB 182 REILGPADNFEQGMWLLAPITVYSQOTRGLLGCITTSLTGRDKQVGEVQVSTAQ 241  
 DB KEILGPADNFEQGMWLLAPITVYSQOTRGLLGCITTSLTGRDKQVGEVQVSTAQ 1067  
 QY 242 SFLATCVNGVCWVYFHAGSKTLAGPKGPTTOMVTWDDQLVGMQAPRGASMTPTCTGS 301  
 DB SFLATCVNGVCWVYFHAGSKTLAGPKGPTTOMVTWDDQLVGMQAPRGASMTPTCTGS 1127  
 QY 1068 SFLATCVNGVCWVYFHAGSKTLAGPKGPTTOMVTWDDQLVGMQAPRGASMTPTCTGS 1127  
 DB SDLYLVTRHADVPVRRRGRSGSLSPRPISYLKSSGGPILCPGSHVVGIFRAAVCTR 1187  
 QY 362 GVAKAVDPIPVESMETTMR 380  
 DB 1188 GVAKAVDPIPVESMETTMR 1206

RESULT 7  
 AAR40223  
 ID AAR40223 standard; protein; 768 AA.  
 XX AAR40223;  
 AC 21-FEB-1994 (first entry)  
 DT 21-FEB-1994 (first entry)  
 DE Recombinant hepatitis C virus genomic protein.  
 XX Hepatitis; HCV; virus; screening; antiviral drugs.  
 KM Hepatitis C virus.  
 OS

Key Location/Qualifiers  
 FH 10  
 FT Misc-difference /note= "UUA encodes Ile."  
 FT Misc-difference 81..82  
 FT /note= "Nucleotide sequence encodes another Gly"  
 FT Duplication 528..768  
 FT /note= "Duplication of 241 amino acids at start of  
 FT protein sequence."  
 FT

```

FT      Misc-difference 537
XX      /note= "UUA encodes Ile."
XX      JP05192160-A.
XX
XX      03-AUG-1993.
XX
XX      20-JAN-1992; 92JP-00028833.
XX
XX      20-JAN-1992; 92JP-00028833.
XX
XX      (BANY ) BANYU PHARM CO LTD.
XX
XX      WPI; 1993-277474/35.
XX
XX      N-PSDB; AAQ48215.
XX
XX      Hepatitis C virus genomic RNA, cDNA and polypeptide - used for screening
XX      hepatitis C virus-specific antiviral drugs.
XX
XX      Claim 9; Page 4-6; 14pp; Japanese.
XX
XX      The protein is useful in the screening of HCV-specific antiviral drugs.
XX      HCV cDNA was cloned from plasma. Plasmids pSR3241 and pSR2541 were
XX      prepared using the cDNA and plasmid pSR3241 was used to transform a COS-1
XX      cell.
XX
XX      Sequence 768 AA.
XX
XX
XX      Query Match          95.0%; Score 1888; DB 2; Length 768;
XX      Best Local Similarity 93.2%; Pred. No. 1.8e-174;
XX      Matches 354; Conservative 13; Mismatches 13; Indels 0; Gaps 0;
XX
XX      1 ALTLSPYKVLARLIMWLQYITLTVKHAHLQWIPPLNVRGRDAIILTCVHPELIF 60
XX      88 ALTLSPYKVLARLIMWLQYITLTVKHAHLQWIPPLNVRGRDAIILTCVHPELIF 147
XX      61 DITKLALIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 120
XX      148 DITKLALIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 207
XX      121 TVYVDHLTPLODMANAGLRDLAVAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARR 180
XX      208 TVYVDHLTPLODMANAGLRDLAVAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARR 267
XX      181 GREILIGPADNFEQGWRLAPITAYSQQTRGLGCIITSLGRDNQVGEVQVSTAT 240
XX      268 GREILIGPADNFEQGWRLAPITAYSQQTRGLGCIITSLGRDNQVGEVQVSTAT 327
XX      241 QSFPLATCVNGVCMVTFHGAAGSKTLGAPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCG 300
XX      328 QSFPLATCVNGVCMVTFHGAAGSKTLGAPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCG 387
XX      301 SSDLIVLTHADVIPTRRRGRSGSLSPRPVSYLKSGSGGPLCPGSHAVGIFRAAVCT 360
XX      388 SSDLIVLTHADVIPTRRRGRSGSLSPRPVSYLKSGSGGPLCPGSHAVGIFRAAVCT 447
XX
XX      361 RGVAKAVDFIPIVESMETTMR 380
XX      448 RGVAKAVDFIPIVESMETTMR 467
XX
XX
XX      RESULT 8
XX      ID ABG30601 standard; protein; 2201 AA.
XX      AC ABG30601;
XX      XX
XX      21-OCT-2002 (first entry)
XX      Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
XX      Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX      cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutuin.
XX

```

```

XX      Hepatitis C virus.
XX      OS Synthetic.
XX
XX      Key Location/Qualifiers
XX      FT Misc-difference 882
XX      FT /label= Arg, Lys
XX      FT Misc-difference 2183
XX      /note= "Wild type Met substituted by Thr"
XX
XX      WO200252015-A2.
XX
XX      04-JUL-2002.
XX
XX      20-DEC-2001; 2001WO-CR001843.
XX
XX      22-DEC-2000; 2000US-0257857P.
XX
XX      (BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.
XX
XX      Kukolj G, Pause A;
XX
XX      WPI; 2002-575382/61.
XX
XX      New self-replicating RNA molecules from Hepatitis C virus (HCV), which
XX      possess enhanced transduction or replication efficiency, useful for
XX      evaluating potential inhibitors of HCV replication.
XX
XX      Claim 3; Page: 140pp; English.
XX
XX      The invention describes a self-replicating hepatitis C virus (HCV)
XX      polynucleotide molecule comprising a 5'-non translated region (NTR),
XX      where guanine at position 1 is substituted for adenine, a HCV polypeptide
XX      region coding for a HCV polypeptide; and a 3'-NTR region. The self-
XX      replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
XX      potential inhibitors of HCV replication. The HCV RNA molecule is also
XX      useful for efficiently establishing cell culture replication. The self-
XX      replicating polynucleotide molecule contains a 5'-NTR, where G at
XX      position 1 is substituted for A, and therefore provides an alternative to
XX      existing systems comprising a self-replicating HCV RNA molecule that, in
XX      conjunction with mutations in the HCV non-structural region, such as the
XX      G(2042)/C/R mutations, transduces and/or replicates with greater
XX      efficiency. This amino acid sequence represents a mutant of the hepatitis
XX      C virus replicon ABGK12 and contains the viral protease NS2/3, protease
XX      CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
XX      This sequence does not appear in the specification but has been created
XX      from the wild type sequence shown in ABG30580 using information given in
XX      the claims of the invention
XX
XX      Sequence 2201 AA.
XX
XX
XX      Query Match          95.0%; Score 1887; DB 5; Length 2201;
XX      Best Local Similarity 93.4%; Pred. No. 9.9e-174;
XX      Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
XX
XX      2 LITLSPYKVLARLIMWLQYITLTVKHAHLQWIPPLNVRGRDAIILTCVHPELIF 61
XX      19 LITLSPYKVLARLIMWLQYITLTVKHAHLQWIPPLNVRGRDAIILTCVHPELIF 78
XX      62 IYKLLAIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 121
XX      79 IYKLLAIFGPMVQAGITKVPYFVRAQGLIRACMLVRKAAAGHYVQMAFMKLAALTG 138
XX      122 TVYVDHLTPLODMANAGLRDLAVAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARRG 181
XX      139 TVYVDHLTPLODMANAGLRDLAVAVEPVIFSDMEVKIITWGAADTAACGDIISGLPVSARRG 198
XX      182 REILIGPADNFEQGWRLAPITAYSQQTRGLGCIITSLGRDNQVGEVQVSTATQ 241
XX      199 REILIGPADNFEQGWRLAPITAYSQQTRGLGCIITSLGRDNQVGEVQVSTATQ 258
XX      242 SFLATCVNGVCMVTFHGAAGSKTLGAPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCG 301
XX

```



DB 259 SFLATCVGVCMTVYHGAGSKTLGAPKGPITOMYTNVDOLVGMQAPPGARSMTPTCTGS 318  
 QY 302 SDLYLVTRHADVIPIVRRRDSRGSILSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
 DB 319 SDLYLVTRHADVIPIVRRRDSRGSILSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 378  
 QY 362 GYAKAVDFPIVPSMETTMR 380  
 DB 379 GYAKAVDFPIVPSMETTMR 397

RESULT 9  
 ABG30591  
 ID ABG30591 standard; protein; 2201 AA.  
 XX ABG30591;  
 AC 21-OCT-2002 (first entry)  
 XX  
 DT Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 DE  
 XX Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 751  
 FT /note= "Wild type Ser substituted by Gly"  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 XX  
 FN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV molecule such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 XX  
 SQ Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
 Best Local Similarity 93.4%; Pred. No. 9,9e-174;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLSPYKXVLLARLIMWLOYLITRVEAHLQWIPPLNVRGGRDAIILLTCVHPELIFD 61  
 DB 19 LTLSPYKXVLLARLIMWLOYLITRVEAHLQWIPPLNVRGGRDAIILLTCVHPELIFD 78  
 QY 62 ITKLLAIFGGIMVLOAGITKVPFVRAAGLIRACMLVRAAGHYVQMAFMKLAALTGT 121  
 DB 79 ITKLLAIFGGIMVLOAGITKVPFVRAAGLIRACMLVRAAGHYVQMAFMKLAALTGT 138  
 QY 122 YVVDHLTPLOPMWAGRDILAVAVEPVIIPSDMEVKIITWQADTAACGDIISGLFVSARG 181  
 DB 139 YVVDHLTPLOPMWAGRDILAVAVEPVIIPSDMEVKIITWQADTAACGDIISGLFVSARG 198  
 QY 182 REILGPADNFEQGMRLAPITAYSOOTRGLCITTSITGRDKNVGEVQVSTATQ 241  
 DB 199 REILGPADNFEQGMRLAPITAYSOOTRGLCITTSITGRDKNVGEVQVSTATQ 258  
 QY 242 SFLATCVGVCMTVYHGAGSKTLGAPKGPITOMYTNVDOLVGMQAPPGARSMTPTCTGS 301  
 DB 259 SFLATCVGVCMTVYHGAGSKTLGAPKGPITOMYTNVDOLVGMQAPPGARSMTPTCTGS 318  
 QY 302 SDLYLVTRHADVIPIVRRRDSRGSILSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
 DB 319 SDLYLVTRHADVIPIVRRRDSRGSILSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 378  
 QY 362 GYAKAVDFPIVPSMETTMR 380  
 DB 379 GYAKAVDFPIVPSMETTMR 397

RESULT 10  
 ABG30600  
 ID ABG30600 standard; protein; 2201 AA.  
 XX ABG30600;  
 AC 21-OCT-2002 (first entry)  
 XX  
 DT Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.  
 DE  
 XX Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 FT Misc-difference 1357  
 FT /note= "Wild type Pro substituted by Leu"  
 XX  
 FN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.

XX Claim 3, Page: 140pp, English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polypeptide  
XX region coding for a HCV polypeptide and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence represents a mutant of the hepatitis  
XX C virus replicon ABG30581 and contains the viral protease NS2/3, protease  
XX complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
XX This sequence does not appear in the specification but has been created  
XX from the wild type sequence shown in ABG30580 using information given in  
XX the claims of the invention

XX Sequence 2201 AA:

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLWMLQYLITRVAHQWIPPLNVRGGRDAIILTCVAPPELLFD 61  
DB 19 LTLSPHYKLFARLWMLQYFITRAEHLQWIPPLNVRGGRDAIILTCVAPPELLFD 78  
QY 62 ITKLALIFGLPMLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 121  
DB 79 ITKLALIFGLPMLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 138  
QY 122 YTFDHLTPLODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181  
DB 139 YVVDHLTPLODMAHAGLRDLAVAVEPIFSDMETKVIITWGADTAACGDIISGLPVASARG 198  
QY 182 REILGPADNFEQGRRLAPITAYSQQTRGLGCIITSLTRDRNQVEGEVQVNSTAQ 241  
DB 199 REILGPADNFEQGRRLAPITAYSQQTRGLGCIITSLTRDRNQVEGEVQVNSTAQ 258  
QY 242 SPLATCNGVCWTVFHGAGSKTLGPKGPIITQWYTNVDDLVGMQAPPARASLTPTCTGS 301  
DB 259 SPLATCNGVCWTVFHGAGSKTLGPKGPIITQWYTNVDDLVGMQAPPARASLTPTCTGS 318  
QY 302 SDLYLTVRHADVIIVRRGDSRGLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
DB 319 SDLYLTVRHADVIIVRRGDSRGLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 378  
QY 362 GVAKAVDFIVESMETTMR 380  
DB 379 GVAKAVDFIVESMETTMR 397

RESULT 11

ABG30581  
ID ABG30581 standard; protein; 2201 AA.

XX ABG30581;

XX 21-OCT-2002 (first entry)

XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.

XX Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor;

XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.

XX Hepatitis C virus.  
XX  
XX WO200252015-A2.

XX 04-JUL-2002.

XX 20-DEC-2001; 2001WO-CA001843.

XX 22-DEC-2000; 2000US-0257857P.

XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;

XX WPI: 2002-575382/61.

XX N-PSDB: ABR88573.

XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
XX possess enhanced transduction or replication efficiency, useful for  
XX evaluating potential inhibitors of HCV replication.

XX Disclosure: Page 49-58; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
XX polynucleotide molecule comprising a 5'-non translated region (NTR),  
XX where guanine at position 1 is substituted for adenine, a HCV polypeptide  
XX region coding for a HCV polypeptide and a 3'-NTR region. The self-  
XX replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
XX potential inhibitors of HCV replication. The HCV RNA molecule is also  
XX useful for efficiently establishing cell culture replication. The self-  
XX replicating polynucleotide molecule contains a 5'-NTR, where G at  
XX position 1 is substituted for A, and therefore provides an alternative to  
XX existing systems comprising a self-replicating HCV RNA molecule that, in  
XX conjunction with mutations in the HCV non-structural region, such as the  
XX G(2042)/C/R mutations, transduces and/or replicates with greater  
XX efficiency. This amino acid sequence is encoded by the hepatitis C virus  
XX replicon ABG30581 and contains the viral protease NS2/3, protease complex  
XX NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA:

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLWMLQYLITRVAHQWIPPLNVRGGRDAIILTCVAPPELLFD 61  
DB 19 LTLSPHYKLFARLWMLQYFITRAEHLQWIPPLNVRGGRDAIILTCVAPPELLFD 78  
QY 62 ITKLALIFGLPMLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 121  
DB 79 ITKLALIFGLPMLVLAQGITKVPYFRAOGLIRACMLVRKAAGHYVQMAFKLAALTGT 138  
QY 122 YTFDHLTPLODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181  
DB 139 YVVDHLTPLODMAHAGLRDLAVAVEPIFSDMETKVIITWGADTAACGDIISGLPVASARG 198  
QY 182 REILGPADNFEQGRRLAPITAYSQQTRGLGCIITSLTRDRNQVEGEVQVNSTAQ 241  
DB 199 REILGPADNFEQGRRLAPITAYSQQTRGLGCIITSLTRDRNQVEGEVQVNSTAQ 258  
QY 242 SPLATCNGVCWTVFHGAGSKTLGPKGPIITQWYTNVDDLVGMQAPPARASLTPTCTGS 301  
DB 259 SPLATCNGVCWTVFHGAGSKTLGPKGPIITQWYTNVDDLVGMQAPPARASLTPTCTGS 318  
QY 302 SDLYLTVRHADVIIVRRGDSRGLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
DB 319 SDLYLTVRHADVIIVRRGDSRGLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 378  
QY 362 GVAKAVDFIVESMETTMR 380  
DB 379 GVAKAVDFIVESMETTMR 397

RESULT 12

ABG30593

ID	ABG30593	standard; protein; 2201 AA.
XX	ABG30593;	
AC		
DT	21-OCT-2002	(first entry)
XX		
DE	Hepatitis C virus NS2/3, NS3/4, NS3 and NSSB mutant #4.	
XX		
KW	Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;	
XX	cell culture replication; NS2/3; NS3/4; NS3; NSSB; mutant; mutein.	
OS	Hepatitis C virus.	
OS	Synthetic.	
FT	Key	Location/Qualifiers
FT	Misc-difference	882
FT		/label= Arg, Lys
FT	Misc-difference	892
FT		/note= "Wild type Leu substituted by Phe"
XX		
PN	WO200252015-A2.	
PD	04-JUL-2002.	
XX		
PF	20-DEC-2001; 2001WO-CA001843.	
XX		
PR	22-DEC-2000; 2000US-0257857P.	
XX		
PA	(BOEH ) BOEHRINGER INGELHEIM CANADA LTD.	
PI	Kukulj G, Pause A;	
XX		
DR	WPI, 2002-575382/61.	
XX		
PT	New self-replicating RNA molecules from Hepatitis C virus (HCV), which	
PT	possess enhanced transduction or replication efficiency, useful for	
PT	evaluating potential inhibitors of HCV replication.	
XX		
PS	Claim 3, Page; 140pp; English.	
XX		
CC	The invention describes a self-replicating hepatitis C virus (HCV)	
CC	polynucleotide molecule comprising a 5'-non translated region (NTR),	
CC	where guanine at position 1 is substituted for adenine, a HCV polyprotein	
CC	region coding for a HCV polyprotein; and a 3'-NTR region. The self-	
CC	replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating	
CC	potential inhibitors of HCV replication. The HCV RNA molecule is also	
CC	useful for efficiently establishing cell culture replication. The self-	
CC	replicating polynucleotide molecule contains a 5'-NTR, where G at	
CC	position 1 is substituted for A, and therefore provides an alternative to	
CC	conjunction systems comprising a self-replicating HCV RNA molecule that, in	
CC	conjunction with mutations in the HCV non-structural region, such as the	
CC	G(2042)C/R mutations, transduces and/or replicates with greater	
CC	efficiency. This amino acid sequence and/or represents a mutant of the hepatitis	
CC	C virus replicon APGX12 and contains the viral protease NS2/3, protease	
CC	complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NSSB. Note:	
CC	This sequence does not appear in the specification but has been created	
CC	from the wild type sequence shown in ABG30580 using information given in	
XX	the claims of the invention	
XX		
SQ	Sequence 2201 AA;	
XX		
QY	Query Match	95.0%; Score 1887; DB 5; Length 2201;
QY	Best Local Similarity	93.4%; Pred. No. 9.9e-174;
QY	Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0	
DB		
DB	19	LTLLSPHYKLFARLLIMWLQYFTRRAHLDQWIPPLNVRGGDAVILITCAIHPELLFT 78
QY	2	LTLLSPHYKLVARLLIMWLQYFTRRAHLDQWIPPLNVRGGDAVILITCAIHPELLFD 61
QY	62	ITTLALIAIGPPLNVLQAGITKYVYFPAQGLPACMLVLRKAAAGHYVQAAEMKLAALTGG 121
DB	79	ITTLALIAIGPPLNVLQAGITKYVYFPAQGLPACMLVLRKAAAGHYVQAAEMKLAALTGT 138

Oy		12	YVVDHLPLDQDMHAGLRDLAAVAEPVIFSDMEVKIITWGADLTACGDIISGLPVSARRG	181
Dd		139	YVVDHLPLDQDMHAGLRDLAAVAEPVFPESEMERKVIITMGADLTACGDIIILGLPVSARRG	198
Oy		182	REILLGPADNFEFGGWELLAPITAVSQOTRGLCCIIITSLTGRDNQVEGEVQVSTATQ	241
Dd		199	REILHGADSIIEGGGWELLAPITAVSQOTRGLCCIIITSLTGRDRNQVEGEVQVSTATQ	258
Oy		242	SFIATCNVCWMTVPHAGSKTLAHPGPIIQWTNVTDQDLVGMQAPPGARSMTPTCGS	301
Dd		259	SFIATCNVCWMTVPHAGSKTLAHPGPIIQWTNVTDQDLVGMQAPPGARSLTPTCGS	318
Oy		302	SDLVLVTRHADLVIPRRRGDSRGSLSPRPVSYLKSGSGGPLCPSGHAVGIFFRAVCTR	361
Dd		319	SDLVLVTRHADLVIPRRRGDSRGSLSPRPVSYLKSGSGGPLCPSGHAVGIFFRAVCTR	378
Oy		362	GVAKANDVPFVESMETTWIR	380
Dd		379	GVAKANDVPFVESMETTWIR	397
RESULT 13				
ID	ABG30582	standard; protein; 2201 AA.		
XX	ABG30582;			
DE	21-OCT-2002	(first entry)		
XX	Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.			
KM	Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;			
OS	cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; nucleic acid			
XX	Hepatitis C virus.			
FT	Key	Location/Qualifiers		
FT	Misc-difference 882	/note= "wild type Lys substituted by Lys or Arg"		
FT	Misc-difference 1233	/note= "wild type Gly substituted by Cys"		
PX	MO200252015-A2.			
PD	04-JUL-2002.			
PF	20-DEC-2001; 2001WO-CAN001843.			
PR	22-DEC-2000; 2000US-0257857P.			
PA	(BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.			
PI	Kukolj G, Pause A;			
XX	WPI; 2002-575382/61.			
DR	N-PDSB; ABK88574.			
PT	New self-replicating RNA molecules from Hepatitis C virus (HCV), which			
PT	possess enhanced transduction or replication efficiency, useful for			
XX	evaluating potential inhibitors of HCV replication.			
XX	Disclosure; Page 59-69; 140pp; English.			
CC	The invention describes a self-replicating hepatitis C virus (HCV)			
CC	polynucleotide molecule comprising a 5'-non translated region (NTR),			
CC	where guanine at position 1 is substituted for adenine, a HCV polypeptide			
CC	region coding for a HCV polypeptide; and a 3'-NTR region. The self-			
CC	replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating			
CC	potential inhibitors of HCV replication. The HCV RNA molecule is also			
CC	useful for efficiently establishing cell culture replication. The self-			
CC	replicating polynucleotide molecule contains a 5'-NTR, where G at			
CC	position 1 is substituted for A, and therefore provides an alternative to			
CC	existing systems comprising a self-replicating HCV RNA molecule that, in			

conjunction with mutations in the HCV non-structural region, such as the G12042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence is encoded by the hepatitis C virus replicon Apgk12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this sequence has been created from replicon Apgk12 shown in ABG30581

Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```
QY 2 LTTSPYKVLARLIMWLQYLITRVAHLQWIPUNVAGRDAILLTCAVHPELIFD 61
DB 19 LTLSPHYKLFARLIMWLQYFTRABAHQWIPUNVAGRDAILLTCAVHPELIF 78
QY 62 ITKLILAFGLMVLQAGITKVPFVFAAGLIRACMLVRRAGGHYQVAFMLAALTGT 121
DB 79 ITKLILAFGLMVLQAGITKVPFVFAAGLIRACMLVRRAGGHYQVAFMLAALTGT 138
QY 122 YVVDHLTPLODMAHAGRDIAVAVEPVTFSDMEVKIITWGDADTAACGDIILGLPVSAARRG 181
DB 139 YVVDHLTPLODMAHAGRDIAVAVEPVTFSDMEVKIITWGDADTAACGDIILGLPVSAARRG 198
QY 182 REILGPDNFEQGMWLLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 241
DB 199 REILGPDNFEQGMWLLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 258
QY 242 SFATCVNGVCMVTFHAGSKTLAGPKGPTTQMTYNVDQDLVGMQAPPGARSLTPTCTGS 301
DB 259 SFATCVNGVCMVTFHAGSKTLAGPKGPTTQMTYNVDQDLVGMQAPPGARSLTPTCTGS 318
QY 302 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
DB 319 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 378
QY 362 GVAKAVDFIVESMETTMR 380
DB 379 GVAKAVDFIVESMETTMR 397
```

#### RESULT 14

ABG30580  
ID ABG30580 standard; protein; 2201 AA.

```
XX AC ABG30580;
XX DT 21-OCT-2002 (first entry)
XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #3.
XX KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX OS Hepatitis C virus.
XX FH Key Location/Qualifiers
XX FT Misc-difference 882 /note= "Encoded by ARG"
XX PN WO200252015-A2.
XX PD 04-JUL-2002.
XX PP 20-DEC-2001; 2001WO-CA001843.
XX PR 22-DEC-2000; 2000US-0257857P.
XX PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX PI Kukulj G, Pause A;
```

DR WPI; 2002-575382/61.  
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.

Disclosure: Page 69-74; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)  
CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
CC useful for efficiently establishing cell culture replication. The self-  
CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
CC position 1 is substituted for A, and therefore provides an alternative to  
CC existing systems comprising a self-replicating HCV RNA molecule that, in  
CC conjunction with mutations in the HCV non-structural region, such as the  
CC G12042)C/R mutations, transduces and/or replicates with greater  
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

Sequence 2201 AA;

Query Match 95.0%; Score 1887; DB 5; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 9.9e-174;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```
QY 2 LTTSPYKVLARLIMWLQYLITRVAHLQWIPUNVAGRDAILLTCAVHPELIFD 61
DB 19 LTLSPHYKLFARLIMWLQYFTRABAHQWIPUNVAGRDAILLTCAVHPELIF 78
QY 62 ITKLILAFGLMVLQAGITKVPFVFAAGLIRACMLVRRAGGHYQVAFMLAALTGT 121
DB 79 ITKLILAFGLMVLQAGITKVPFVFAAGLIRACMLVRRAGGHYQVAFMLAALTGT 138
QY 122 YVVDHLTPLODMAHAGRDIAVAVEPVTFSDMEVKIITWGDADTAACGDIILGLPVSAARRG 181
DB 139 YVVDHLTPLODMAHAGRDIAVAVEPVTFSDMEVKIITWGDADTAACGDIILGLPVSAARRG 198
QY 182 REILGPDNFEQGMWLLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 241
DB 199 REILGPDNFEQGMWLLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVSTATQ 258
QY 242 SFATCVNGVCMVTFHAGSKTLAGPKGPTTQMTYNVDQDLVGMQAPPGARSLTPTCTGS 301
DB 259 SFATCVNGVCMVTFHAGSKTLAGPKGPTTQMTYNVDQDLVGMQAPPGARSLTPTCTGS 318
QY 302 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
DB 319 SDLYLVTRHADVIPVRRGDSRGLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 378
QY 362 GVAKAVDFIVESMETTMR 380
DB 379 GVAKAVDFIVESMETTMR 397
```

#### RESULT 15

ABG30587  
ID ABG30587 standard; protein; 2201 AA.

```
XX AC ABG30587;
XX DT 21-OCT-2002 (first entry)
XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #7.
XX KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
XX KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.
XX OS Hepatitis C virus.
```